sas

# Fraud Analytics with SAS®

## Special Collection


Foreword by
Bart Baesens

The correct bibliographic citation for this manual is as follows: Baesens, Bart. 2017. *Fraud Analytics with SAS®: Special Collection*. Cary, NC: SAS Institute Inc.

**Fraud Analytics with SAS®: Special Collection**

Copyright © 2017, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

September 2017

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.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to **http://support.sas.com/thirdpartylicenses**.

# Table of Contents

# Free SAS® e-Books: Special Collection

In this series, we have carefully curated a collection of papers that introduces and provides context to the various areas of analytics. Topics covered illustrate the power of SAS solutions that are available as tools for data analysis, highlighting a variety of commonly used techniques.


Discover more free SAS e-books!
**support.sas.com/freesasebooks**

sas.com/books
*for additional books and resources.*

sas
THE POWER TO KNOW®

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. © 2017 SAS Institute Inc. All rights reserved. M1673525 US.0817

# About This Book

## What Does This Collection Cover?

Current thinking in fraud detection is moving away from the silo approach and recognizing the need for a more proactive and holistic approach to data and analytics. An isolated event may be flagged as suspicious, but without a complete view of the interplaying relationships, the investigator might ignore it.

SAS software provides many different techniques to monitor in real time and investigate your data, and several groundbreaking papers have been written to demonstrate how to use these techniques. In this free e-collection, we carefully selected a handful of these from recent SAS Global Forum papers to introduce you to the topics and to let you sample what each has to offer.

The following papers are excerpts from the SAS Global Users Group *Proceedings*. For more SUGI and SAS Global Forum *Proceedings*, visit the online versions of the Proceedings.

For many more helpful resources, please visit support.sas.com and sas.com/books.

## We Want to Hear from You

SAS Press books are written *by* SAS users *for* SAS users. We welcome your participation in their development and your feedback on SAS Press books that you are using. Please visit sas.com/books to

- Sign up to review a book
- Request information on how to become a SAS Press author
- Recommend a topic
- Provide feedback on a book

Do you have questions about a SAS Press book that you are reading? Contact the author through saspress@sas.com.

# Foreword

It is estimated that a typical organization loses about 5% of its revenues to fraud each year (www.acfe.com). A detailed characterization of the multifaceted phenomenon of fraud has been provided by Van Vlasselaer as follows:

> *Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms.*

This definition highlights five characteristics that are associated with particular challenges related to developing an analytical fraud detection system. Fraud is uncommon since only a minority of the population of cases typically engages in fraud, which makes it difficult to detect. Moreover, fraudsters will try to blend in to avoid being noticed and to remain obscured by non-fraudsters. This effectively makes fraud imperceptibly concealed because fraudsters aim to hide by carefully considering and planning how to precisely commit fraud without detection. Fraud detection systems improve and learn by example; therefore, the techniques and tricks that fraudsters adopt evolve over time along with or, better, ahead of fraud detection mechanisms. Fraud is often a carefully organized crime as well, meaning that fraudsters often do not operate independently, have allies, and may induce copycats. A final element in the description of fraud provided by Van Vlasselaer indicates the many different types of forms in which fraud occurs. This refers to both the wide set of techniques and approaches used by fraudsters and to the many different settings in which fraud occurs. Some popular examples are credit card fraud, insurance claim fraud, anti-money laundering, identity theft, insurance fraud, corruption, counterfeit, product warranty fraud, telecommunications fraud, click fraud, and tax evasion.
Current thinking in fraud detection is moving away from the silo approach and recognizing that a more proactive and holistic approach to data and analytics is needed. An isolated event may be flagged as suspicious, but without a complete view of the interplaying relationships, investigators may ignore it.

SAS software provides many different techniques to monitor in real time and investigate your data, and several ground breaking papers have been written to demonstrate how to use these techniques. We have carefully selected a handful of these from recent SAS Global Forum papers to introduce you to the topics and let you sample what each has to offer.

Renee Elizabeth Palmer - Money Launderers Beware: Catching You is Just a Point and Click Away

As analysts, we want our scenarios to be spot on to catch the criminals. This paper describes the different ways that the data can be explored to detect anomalous patterns and how to create the scenarios in SAS Visual Scenario Designer; test and tune the scenarios and parameters; and how the alerts will be ported seamlessly into the SAS Anti-Money Laundering alert generation process and available for management in the SAS Enterprise Case Management System.

Danielle Davis, Stephen Boyd, and Ray Ong - Minimizing Fraud Risk Through Dynamic Entity Resolution and Network Analysis

Every day, businesses have to remain vigilant for fraudulent activity, which threatens customers, partners, employees, and financials. Normally, networks of people or groups perpetrate deviant activity. Finding these connections is now made easier for analysts with SAS Visual Investigator. This paper discusses how the network analysis view of SAS Visual Investigator, with all its dynamic visual capabilities, can make the investigative process more informative and efficient.

Kathy Hart and Malcolm Alexander - Alerts Don't Launder Money (or Finance Terrorism) – People Do!
For far too long, anti-money laundering and terrorist financing solutions have forced analysts to wade through oceans of transactions and alerted work items (alerts). Rather than starting with alerts and transactions, starting with a customer-centric view allows your analysts to rapidly triage suspicious activities, prioritize work, and quickly move to investigating the highest risk customer activities. This paper discusses how a customer-centric approach leads to increased analyst efficiency and streamlined investigations.

Gordon Robinson and Ryan Schmiedl - Adding a Workflow to Your Analytics with SAS Visual Investigator
Analytics is only as good as the decisions that it enables. The key to making the right decisions is ensuring that the right information is given to the right people at the right time. This paper walks through an example of using the administrative tools of SAS Visual Investigator to create a ticketing system in response to threats to a business. It shows how SAS Visual Investigator can easily be adapted to meet the changing nature of the threats the business faces.

Lawrie Elder - Counter Radicalization through Investigative Insights and Data Exploitation Using SAS Viya
This paper illustrates how SAS Viya can aid intelligence, homeland security, and law-enforcement agencies in counterterrorism activities by enabling a hub to cross-reference both internally and ingested externally held data and, crucially, operational intelligence gained from normal policing activities.

Susan Trueman - You Imported What? Supporting International Trade with Advanced Analytics
Global trade and more people and freight moving across international borders present border and security agencies with a difficult challenge. This paper shows how the new features in SAS Visual Investigator can help by bringing together disparate data, detecting suspicious activity, presenting analysts with alerts to be triaged, and enabling investigators and intelligence analysts to conduct investigations and make appropriate, data-driven decisions.

Brooke Fortson and Gordon Robinson - Investigating Connections between Disparate Data Sources with SAS Visual Investigator
Bringing together data around seemingly disparate events can be a lengthy process. For example, in 1993, Erin Brockovich began a lengthy manual investigation after discovering a link between elevated clusters of cancer cases and contaminated water in the same area due to the disposal of chemicals from a utility company. This paper describes how to identify cancer clusters in a geographical location and see whether there is any correlation of those clusters to contaminated areas using the map and network functionalities of SAS Visual Investigator.

John Shipway - Fighting Crime in Real Time with SAS Visual Scenario Designer
The strains that modern criminals are placing on financial and government institutions demand new approaches to detecting and fighting crime. Traditional methods of analyzing large data sets on a periodic batch basis are no longer sufficient. In this paper, the author explores the technology architecture, data flow tools, and methodologies that are required to build a solution based on SAS Visual Scenario Designer, enabling organizations to fight crime in real time.

Edwin Rivera, Jim West, and Carl Suplee - Addressing AML Regulatory Pressures by Creating Customer Risk Rating Models with Ordinal Logistic Regression

With increasing regulatory emphasis on using more scientific statistical processes and procedures in the Bank Secrecy Act/Anti-Money Laundering (BSA/AML) compliance space, financial institutions are being pressured to replace their heuristic, rule-based customer risk rating models with well-established, academically supported, statistically based models. This paper compares heuristic, rule-based models and statistically based models and suggests ordinal logistic regression as an effective statistical modeling technique for assessing customer BSA/AML compliance risk.

We hope these selections give you a useful overview of the many tools and techniques that are available to analyze your data and unearth fraud.

Additionally, SAS also features an e-learning course on the topic. This new course shows how learning fraud patterns from historical data can help fight fraud. For more information, visit: Fraud Detection Using Descriptive, Predictive, and Social Network Analytics.

Prof. Dr. Bart Baesens
Faculty of Economics and Business
www.dataminingapps.com

Bart Baesens is an associate professor at KU Leuven (Belgium) and a lecturer at the University of Southampton (United Kingdom), as well as an internationally known data analytics consultant. He is a foremost researcher in the areas of web analytics, customer relationship management, and fraud detection. His findings have been published in well-known international journals including *Machine Learning* and *Management Science*. Baesens is also co-author of the book *Credit Risk Management: Basic Concepts.*

**Paper SAS6342-2016**

# Money Launderers Beware! Catching You is Just a Point and Click Away

Renee Elizabeth Palmer, SAS Institute Inc.

## ABSTRACT

Wouldn't it be fantastic to develop and tune scenarios in SAS® Visual Scenario Designer and then smoothly incorporate them into your SAS® Anti-Money Laundering solution with just a few clicks of your mouse? Well, now there is a way. SAS Visual Scenario Designer is the first data-driven solution for interactive rule and scenario authoring, testing, and validation. It facilitates exploration, visualization, detection, rule writing, auditing, and parameter tuning to reduce false positives; and all of these tasks are performed using point and click. No SAS coding skills required! Using the approach detailed in this white paper, we demonstrate how you can seamlessly port your SAS Visual Scenario Designer scenarios into the SAS Anti-Money Laundering solution. Rewriting the SAS Visual Scenario Designer scenarios in Base SAS is no longer required! Furthermore, the SAS Visual Scenario Designer scenarios are executed on the lightning speed SAS LASR Analytic Server, reducing the time of the SAS Anti-Money Laundering scenario nightly batch run. The results of both the traditional SAS Anti-Money Laundering alerts and SAS Visual Scenario Designer alerts are combined and available for display on the SAS Enterprise Case Management interface. This white describes the different ways that the data can be explored to detect anomalous patterns and the three mechanisms to translate these patterns into rules. It also documents how to create the scenarios in SAS Visual Scenario Designer; test and tune the scenarios and parameters; and how the alerts will be ported seamlessly into SAS Anti-Money Laundering alert generation process and available for management in the SAS Enterprise Case Management System.

## INTRODUCTION

We want to annihilate the bad guys: the terrorists, the drug cartels, the human traffickers, the thieves. But, how can this be accomplished when it is extremely time consuming to develop the anti-money laundering scenarios? When the pool of domain experts is limited due to scarcity of programming skills? When faced with the challenge of tuning scenario parameters to reduce the number of false positives while not missing the outliers? Wouldn't it be fantastic to develop and tune scenarios through an easy-to-use graphical interface that lets you create scenarios with just a few clicks of the mouse? Well, now there is a way. SAS Visual Scenario Designer allows you to create and tune scenarios and then smoothly incorporate them into your SAS Anti-Money Laundering solution.

As analysts, we want our scenarios to be spot on to catch the criminals. Nevertheless, many banks struggle to keep up with the ever changing compliance laws and, as a result, are faced with the possibility of astronomical fines. Banking management cringe when they read news such as:

- *The Financial Conduct Authority fined Barclays GBP72 million for "failing to minimize risk".*
- *FinCEN Fines Oppenheimer & Co. Inc. $20 Million for Continued Anti–Money Laundering Shortfalls.*
- *The Wall Street Journal reports that Citigroup Inc expects to spend $2.7 billion on legal charges tied to continuing investigations including probes into how the bank has handled anti-money laundering compliances.*

How can scenarios be developed efficiently and tuned accurately; incorporated into the SAS Anti-Money Laundering alert generation process and front end Case Management interface; while harnessing the in-memory parallel and distributed computing capacity of the SAS LASR server? Now there is a solution. SAS Visual Scenario Designer integrated with SAS Anti-Money Laundering.

With the point and click interface backed by the in-memory computing power of SAS Visual Scenario Designer, you can perform data preparation, scenario rule creation, tuning, and deployment. The audit trail generated by the scenario tuning process will satisfy the auditors and reduce the probability of being slapped with costly non-compliance fines. The scenarios can be seamlessly incorporated into the SAS Anti-Money Laundering solution and used by alert generation process to create alerts. The alerts generated by the scenarios developed in SAS Visual Scenario Designer can be combined with the alerts generated from the traditional SAS Anti-Money Laundering alert generation process. The SAS Enterprise Case Management interface can then be used to view and triage alerts, analyze subjects, incidents, cases, reports, and e-files giving the investigators the tools they need to catch the criminals.

## SAS VISUAL SCENARIO DESIGNER OVERVIEW

Criminals are constantly altering how they perpetrate their crimes in an effort to stay one step ahead of the law. Consequently, it is up to the analyst to scrutinize the prolific amounts of data to find these ever changing trends of suspicious behavior. In addition, the parameters must be fine-tuned so that false positives are minimized while not allowing outliers to slip through unnoticed. SAS Visual Scenario Designer has the tools to do this.

### DATA EXPLORATION

The Exploration Scenario Window provides visual design tools that facilitate the ability to explore the data and discover and establish conditions that aid in identifying anomalous patterns of nefarious activities. This tool is beneficial when you are searching for trends, rather than having a predefined criteria.

For more complex exploration SAS Visual Scenario Designer offers the capacity to create decision trees using a straightforward, yet comprehensive drag and drop feature. The decision tree's structure enables you to determine the most effective rules and corresponding thresholds within the data. As before, this tool is beneficial when you are searching for trends, rather than having predefined criteria for the scenario.

You may be tasked to create scenarios with concrete, predefined rules. SAS Visual Scenario Designer has a custom tool for this as well. Decision tables enable you to drag and drop the relevant columns and easily create the associated rules and actions that comprise the scenario.

Once you are satisfied with the results of your data exploration, you can then finalize the actual scenario creation. An overview of the required steps are detailed in the next section.

## SCENARIO RULE DEVELOPMENT

### SAS VISUAL SCENARIO DESIGNER COMPONENTS

SAS Visual Scenario Designer consists of four main components: windows, scenarios, deployments, and simulations. In order to create and tune scenarios and integrate them into the SAS Anti-Money Laundering solution, you use all four of these components.

### DATA PREPARATION USING SAS VISUAL SCENARIO DESIGNER WINDOWS

A SAS Visual Scenario Designer window is a palette used to create a data table. The SAS Visual Scenario Designer window provides the capacity to join one or more LASR tables together and gives the ability to select the columns from these tables to be included in the output table that is generated. Custom columns can be created and various types of aggregations can be performed. In addition, date-based lookbacks and filters can be applied at both the window and column levels.

Each entity type requires its own package, referred to as a deployment. Examples of entity types are *Account*, *Customer* and *Bank*. For each entity, you must construct two different windows as part of the scenario creation process. If you require the associated triggering transactions, an additional window needs to be created.

## INTRODUCTION TO SCENARIO DEVELOPMENT

To give you an overview of the process, we use two different *Account*-based scenarios as examples. Both of these scenarios have been simplified so that the explanation is more tangible.

***SAS10016*** scenario looks for a pattern of cash deposits, each less than $10,000, within a short timeframe. This scenario applies exclusively to personal bank accounts and analyzes only successful transactions. If the sum of the deposits and the number of different branches and deposits made exceed their thresholds, then the alert indicator is set to 1 for this account. The scenario requires three different threshold parameters:

- *Number of branches*
- *Number of transactions*
- *Total amount of transactions*

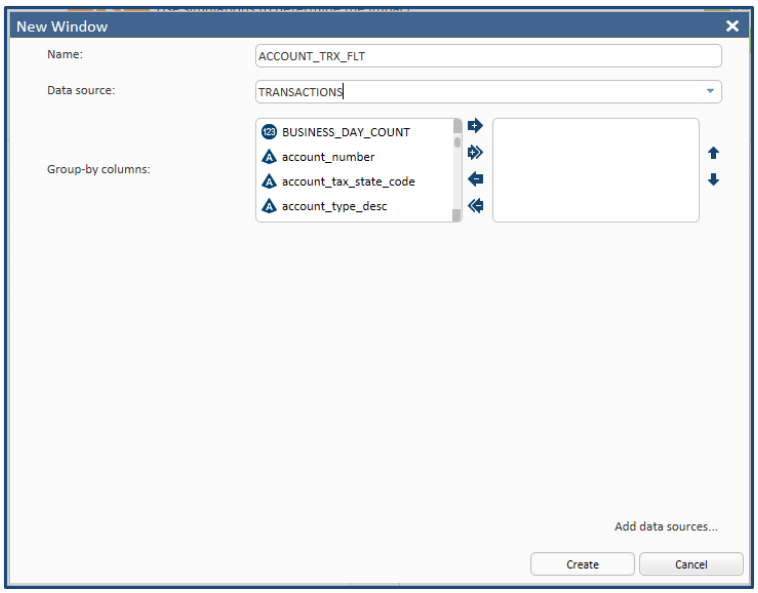
***SAS10019*** scenario calculates the total amount of an account's wire-in and wire-out transactions for the day. If the wires out divided by the wires in exceeds the percentage threshold, then the alert indicator is set to 1 for this account. This scenario also applies exclusively to personal bank accounts and analyzes only successful transactions.

- *Total amount in*
- *Total amount out*

The process of creating both of the scenarios is documented in detail in *Chapter 5 – Visual Scenario Designer* in the *SAS Anti-Money Laundering 6.3: Scenario Administration User's Guide, Second Edition*.

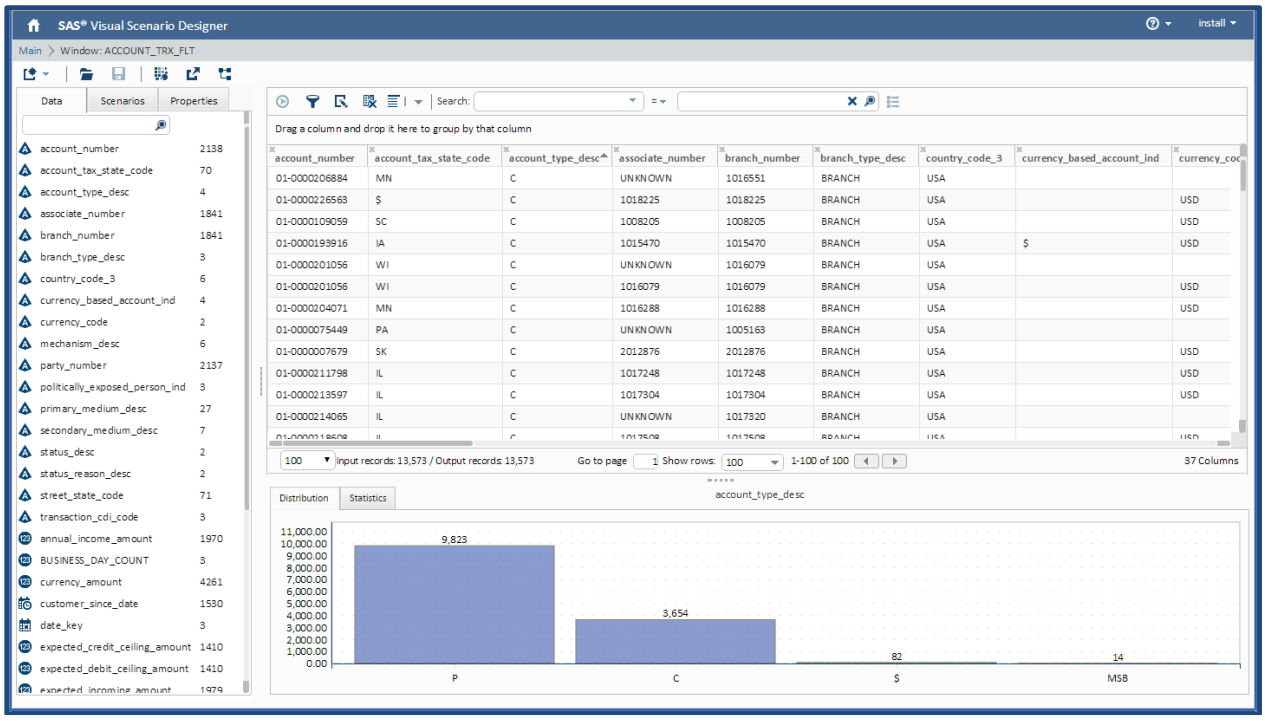
## WINDOW 1 – TRANSFORMATION

Select the appropriate table or tables from the list that have been loaded into LASR. Grouping should not be done within this window, therefore, there is only one row per transaction as shown from the screenshot below.



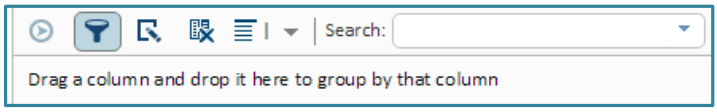

**Display 1. New Window #1**

The output table generated by this window is created by clicking on the available and relevant columns in the data tab and dragging them into the window interface. You can view the data distribution by clicking on any of the columns in the window and the corresponding distribution is displayed in the *Distribution* tab. In addition, you can view statistics about a particular numeric column by clicking on the column and then clicking on the *Statistics* tab. An example of a transaction based window is shown in the screenshot below.



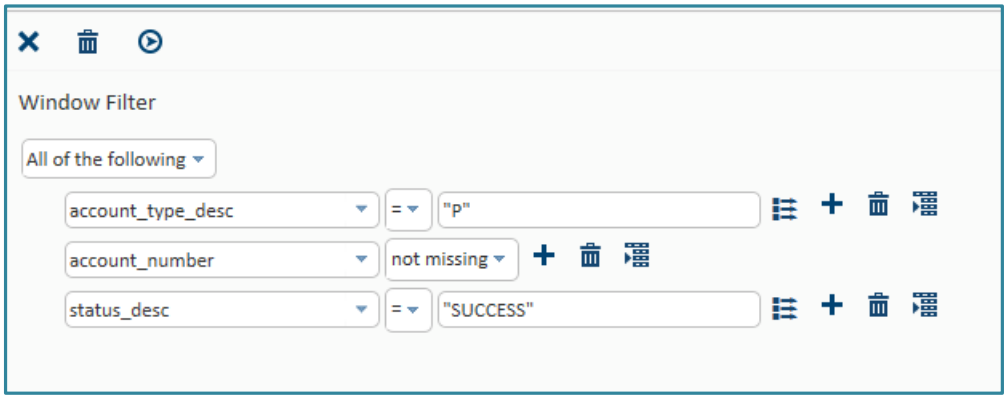

**Display 2. Window View**

SAS Visual Scenario Designer provides the capability to filter data at both the window level and the column level. In this case, both the scenarios we create only analyze successful transactions where the Account type is Personal, '*P*' and where the account number is not missing. We can filter the data channeled into the initial window. This can be done by clicking on the filter icon, which is highlighted in the screen shot below.




**Display 3. Filter**

You can easily enter the filter criteria using the point and click interface provided as shown in the screenshot below.




**Display 4. Window level Filter**

The two sample scenarios require rules to be applied at both the transaction level and at the aggregated Account level. Therefore, scenario rules are created in both window 1 and window 2.

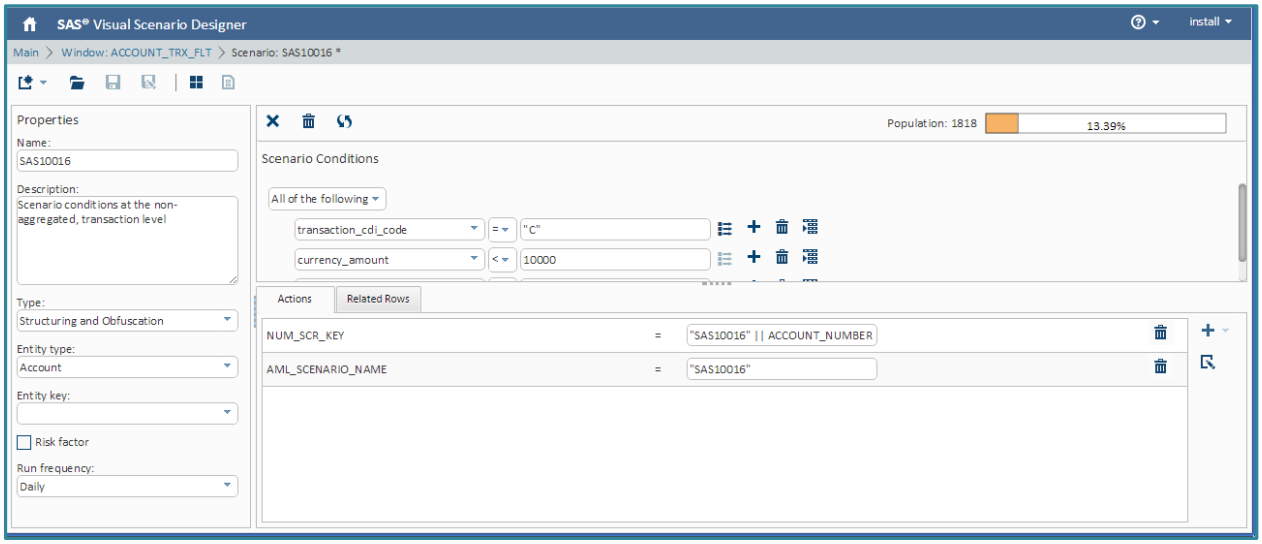
Scenarios are comprised of conditions and actions. In order to create the scenario rule(s), you need to create the condition(s) that define the rule(s) as well as the corresponding action(s) that occur as a result of the condition(s) of the rule(s) being met. An action dictates the name of the new column that is added to the output window along with the value it is populated with if the condition is met. This value could be a numeric value or a string value.

As part of the first window, you are required to create the scenario conditions and actions listed in the bullets below. If the conditions are met, the columns noted in the *Actions* tab are populated with their corresponding values. In this case, *NUM_SCR_KEY* is populated with *SAS10016 || ACCOUNT_NUMBER* and *AML_SCENARIO_NAME* with SAS10016. The *NUM_SRC_KEY* is used to determine which transactions are aggregated and analyzed in the input data table of the next window. The *AML_SCENARIO_NAME* is used to identify the triggering transaction in a third window, should they be required. The use of these two columns is discussed in detail later in this paper.

**SAS10016**

- Conditions applied at the transaction level:
  - *Transaction_cdi_code = "C"*
  - *Currency_amount <= 10000*

- Actions when conditions are met:
  - ***AML_SCENARIO_NAME**:* This action populates the *AML_SCENARIO_NAME* column with the string *SAS10016*, when the corresponding conditions are met.
  - ***NUM_SRC_KEY***: The action populates the *NUM_SRC_KEY* column with the scenario name concatenated with the entity number, when the corresponding condition is met. The column will be populated with *SAS10016* || *Account_number*.

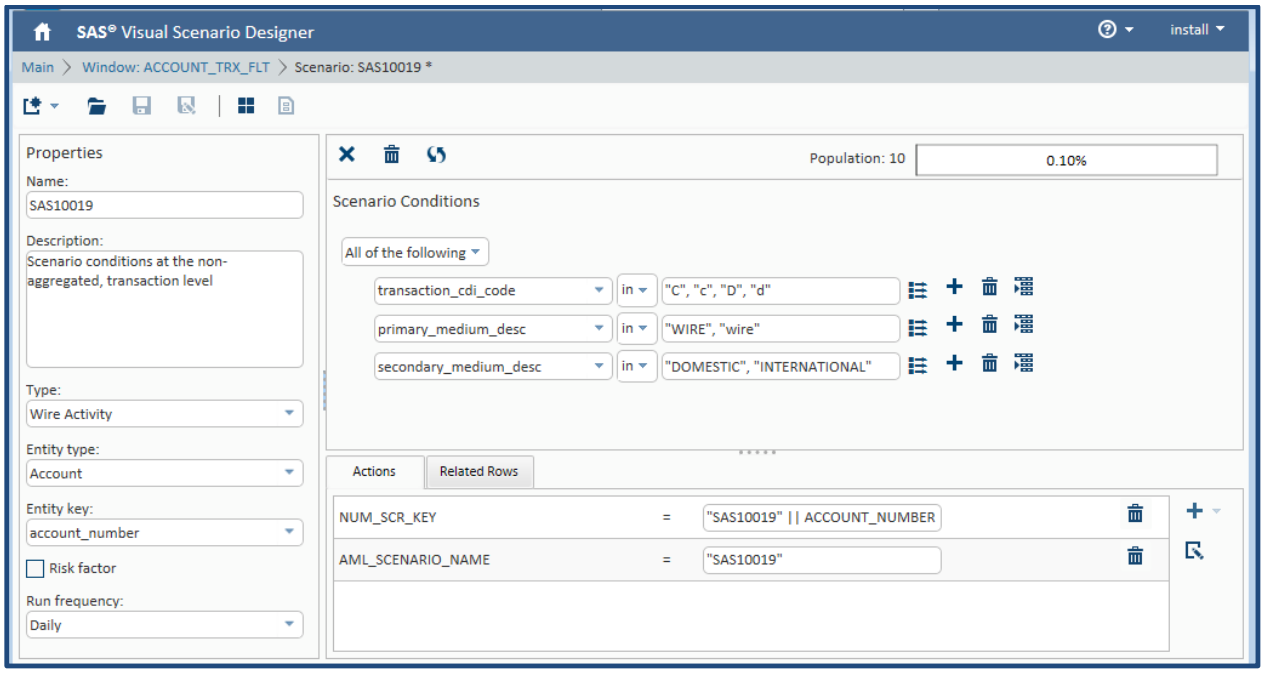



**Display 5. Scenario Conditions – SAS10016**

**SAS10019**

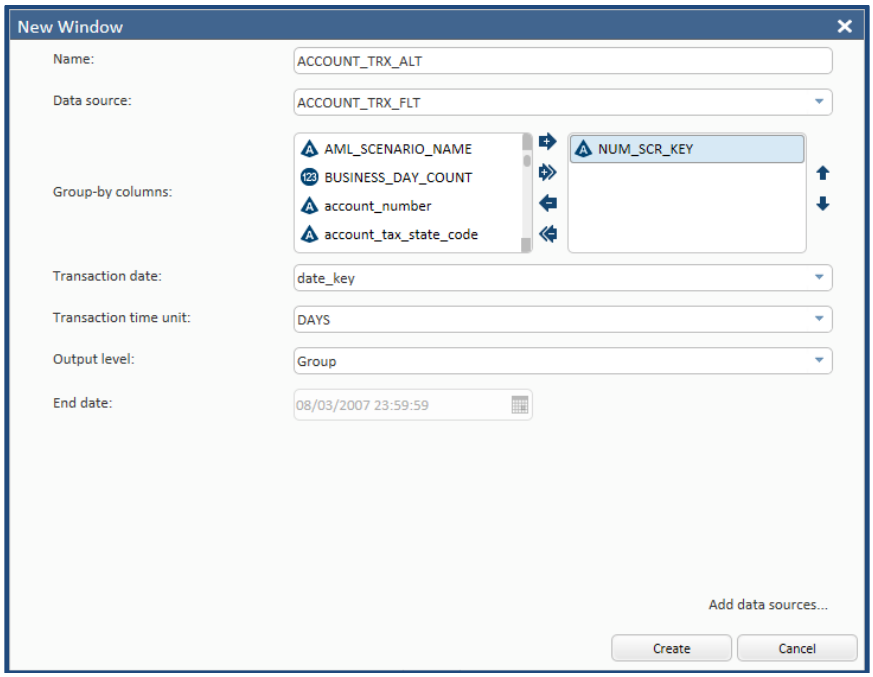
- Conditions applied at the transaction level:
  - *Transaction_cdi_code* in *"C", "c", "D", "d"*
  - *Primary_medium_desc in "WIRE", "wire"*
  - *Second_medium_desc in "DOMESTIC", "INTERNATIONAL"*

- Actions when conditions are met:
  - ***AML_SCENARIO_NAME***: The action populates the *AML_SCENARIO_NAME* column with the string *SAS10019*, when the corresponding condition is met.
  - ***NUM_SRC_KEY**:* The action populates the *NUM_SRC_KEY* column with the scenario name concatenated with the entity number, when the corresponding condition is met. The column will be populated with *SAS10019* || *Account_number*.




**Display 6. Scenario Conditions – SAS10019**

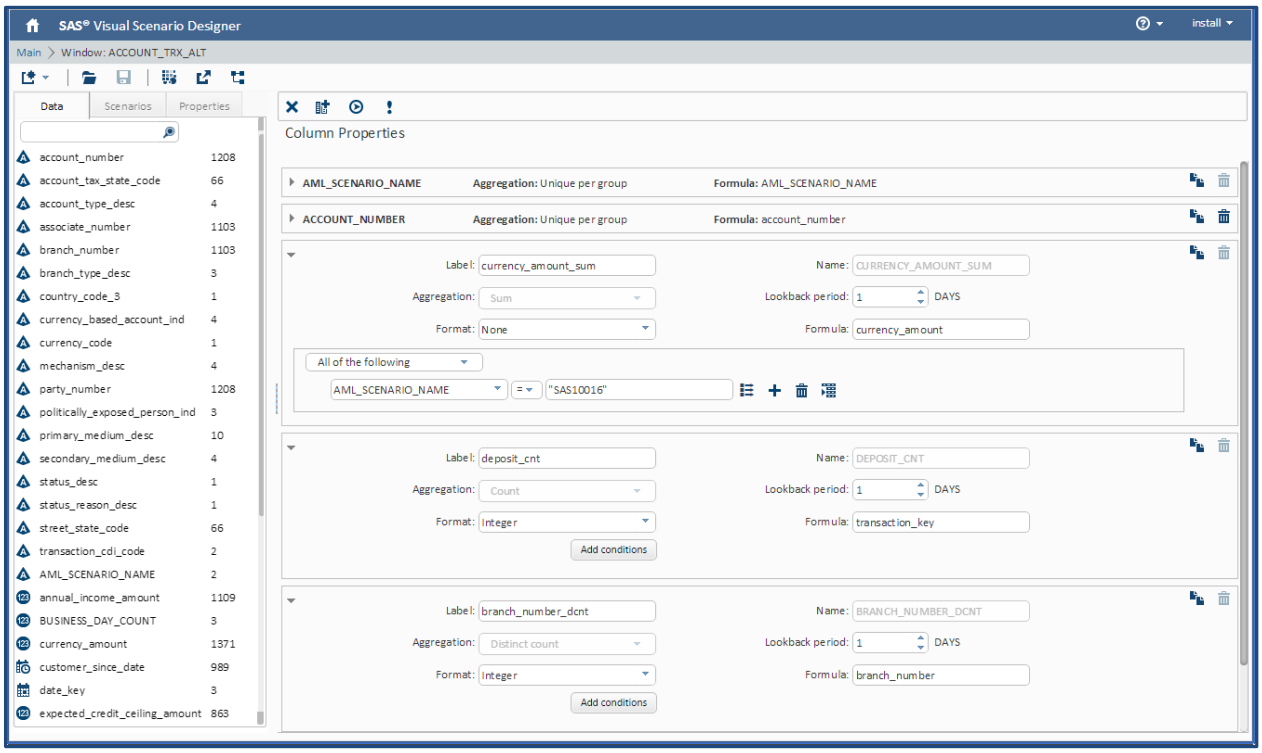
When configuring the output of the second window, select *Group by NUM_SRC_KEY* as shown in the screen shot below. This causes the output of the second window to be the aggregated data from the output of the first window. The aggregation is calculated at the scenario-entity number level stored in the *NUM_SRC_KEY* column, in this case *SAS10016* || *Account* and *SAS10019* || *Account* described earlier.




**Display 7. New Window #2**

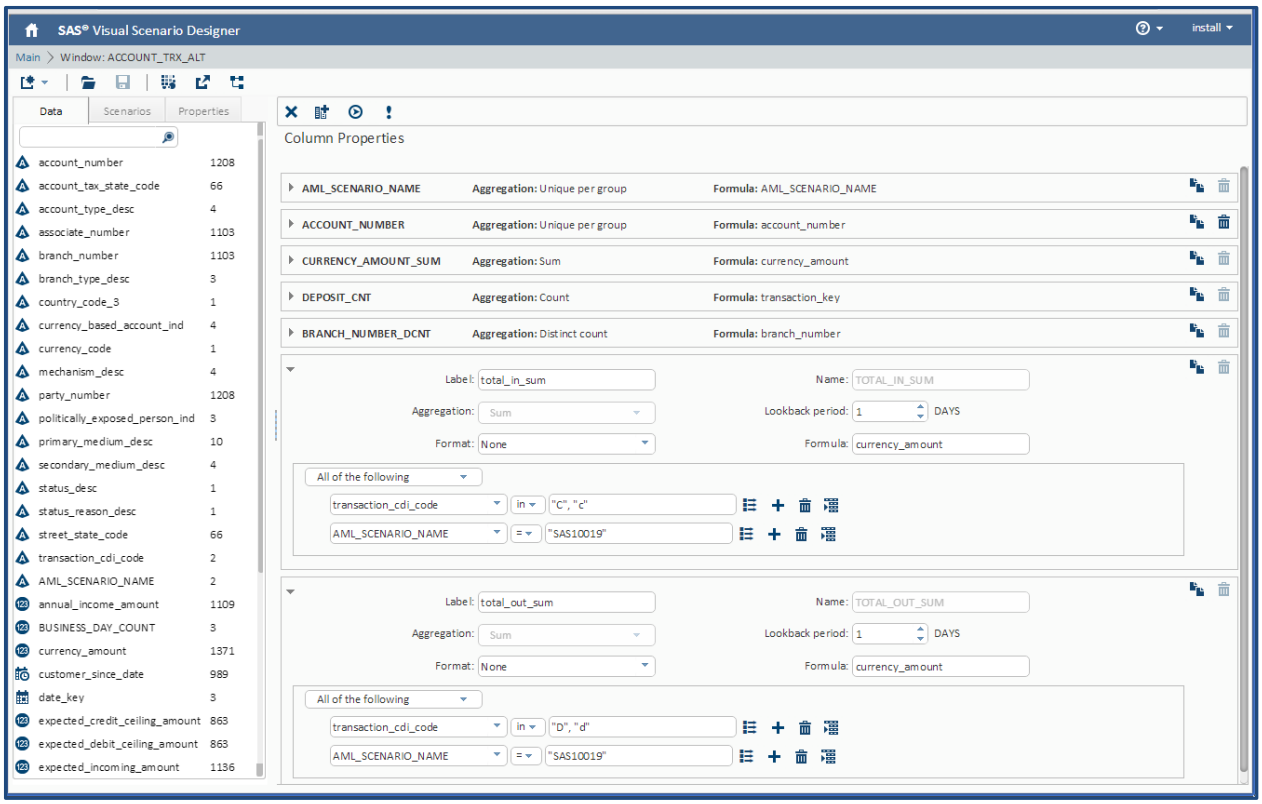
Since the input data from the first window contains transactions that match the conditions dictated by more than one scenario, we must devise a way to funnel the transactions to the second window's scenarios appropriately. We can accomplish this by creating a new column for each rule and applying a filter. By doing so, that column receives only the transactions relevant for the corresponding scenario as rule gets applied and the column gets aggregated. For scenario *SAS10016*, we need the aggregated amount of the transactions, the number of branches where the deposits were made, and the number of deposits. To satisfy the rules, we create three columns for this scenario. Each column applies one of the scenario conditions along with a filter where *AML_SCENARIO = SAS10016*.

As you can see in the screen shot below, filtering on the scenario prior to the aggregation process can be accomplished by adding a condition to the relevant columns. In this case, Scenario *SAS10016* has three columns that require aggregation: *currency_amount_sum*, *deposit_cnt*, and *branch_number_dcnt*. By applying this filter, their corresponding results include only the transactions that met the criteria for *SAS10016* from the previous window.




**Display 8. Column Properties – SAS10016**

For scenario *SAS10019*, we want to aggregate the currency amount of type *Wire in* and *Wire out*. We have already filtered data from the first window such that it contains only wire based transactions for *SAS10019*. To calculate the total *Wire in* amount, we create a column *total_in_sum* that aggregates the *currency_amount* only when the transaction type is "*C*", indicating credit activities. To calculate the total *Wire out* amount, we create a column *total_out_sum*, where the transaction type is "*D*", indicating debit activities. As you can see from the screen shot below, two new columns have been created with the corresponding conditions applied. We can see that we are only aggregating the transactions for *SAS10019* that met the criteria for *SAS10019* in the previous window.




**Display 9. Column Properties – SAS10019**

Now that each scenario has aggregated data matching the criteria for the scenario, we can take the second step in the scenario creation process. The conditions are created by dragging over the columns of interest and setting their corresponding thresholds.

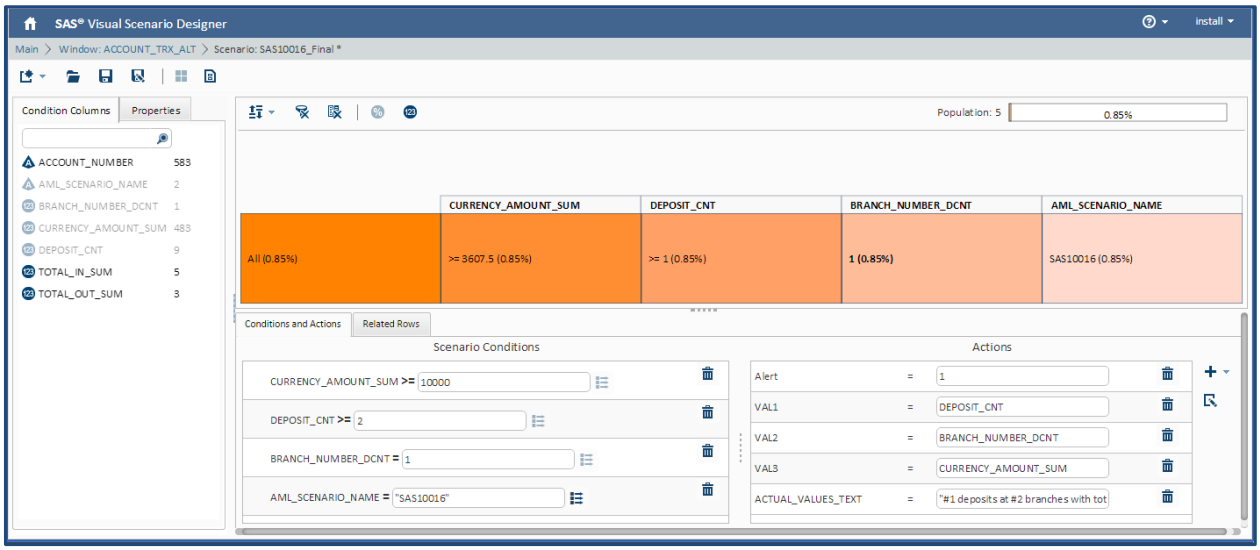
Next, you need to create the actions that should occur when all of the condition are met. At a minimum, you need to create *Alert* and *Actual_Values_Text* actions. *Alert* is set to 1 if the conditions are met, otherwise, it is 0.

- **Alert** = 1 : this is used to indicate whether or not the alert was generated
- **ACTUAL_VALUES_TEXT** = "*#1 deposits at #2 branches with total amount $#3*"

The *Actual_Values_Text* is the message that gets displayed in the SAS Anti-Money Laundering interface. You have the freedom to create any message you like, however, best practices recommend that it describes the scenario and includes the actual triggering values. For each triggering value, you need to add a *#n* to indicate which parameter needs to be included, where *#1* would be a place holder for the first triggering value, *#2*, the second and so forth. An example of this would be: *#1 deposits at #2 branches with total amounts $#3'.*

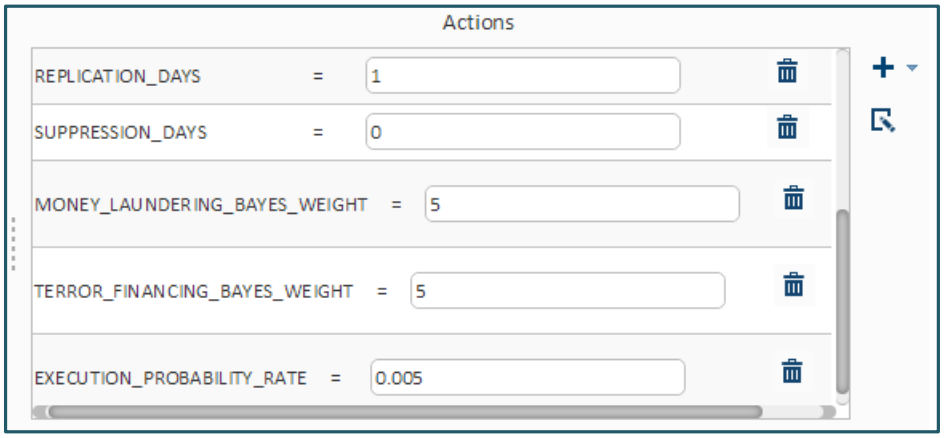
You are required to create an action for each parameter that is embedded in the *Actual_Values_Text*. Each action contains the actual triggering transaction value. Based on the *Actual_Values_Text* example provided, we are required to create three additional parameters. Use the naming convention of *VAL#* for this.

- **VAL1** = DEPOSIT_CNT : aggregated sum of number of deposits
- **VAL2** = BRANCH_NUMBER_DCNT : aggregated sum of branches
- **VAL3** = CURRENCY_AMOUNT_SUM : aggregated sum of deposits



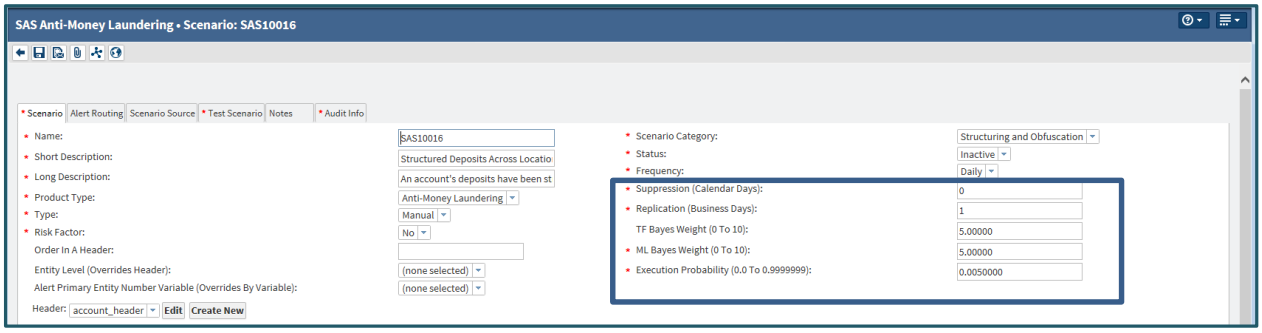

**Display 20. Scenario Conditions and Actions – SAS10016**

There are other parameters that can be set that affect the alert generation and the scoring process. An action can be created for any or all of these, otherwise, the default value is used.




**Display 31. Optional Actions**

These values are displayed in the SAS Anti-Money Laundering Scenario Administrator below.



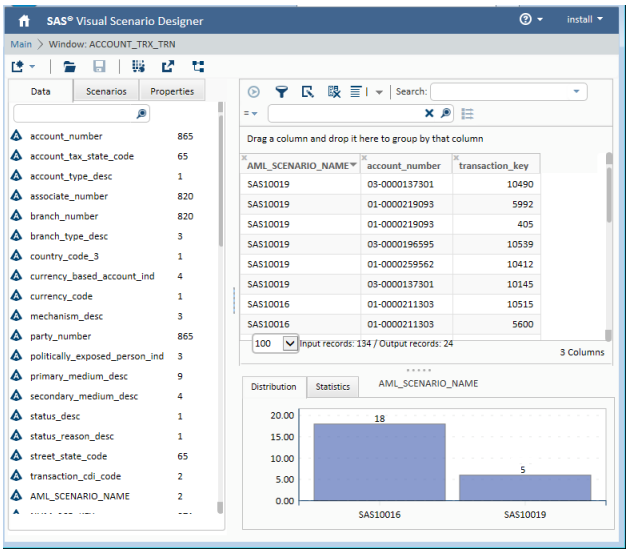

**Display 42. SAS Anti-Money Laundering Scenario Screen**

A complete list of these optional rules are as follows:

- **REPLICATION_DAYS**:
- Usage: Number of days to replicate corresponding transactions
    - Default value: 1
    - Usage: Number of days to suppress alerts for this scenario
    - Default: 0
- **MONEY_LAUNDERING_BAYES_WEIGHT_NUMBERS**
    - This value is used during alert generation process
    - Default: 5
    - Range: 0-10
- **TERROR_FINANCING_BAYES_WEIGH**
    - This value is used during alert generation process
    - Default: 5
    - Range: 0-10
- **EXECUTING_PROBABILITY_RATE**
    - Default: 0.005
    - Range: 0-0.9999999

### Triggering Transactions Window

The investigator may want to view the transactions that triggered the alert, as well as including them as part of a case. The third window contains the triggering transactions for the alert. This window is created by joining the first two windows together by the *NUM_SRC_KEY* column.




**Display 53. Triggering Transaction Window**

Once the windows and the scenarios are created, you are required to create a deployment package for each entity type. In this example, we need to create only one deployment. The deployment contains the windows and scenarios that you created in the previous paragraphs. You can enable and disable the windows and scenarios within the deployment.

The deployment also provides a simulation window. Simulations provides the capacity to execute your scenarios on the fly over data contained in the window to tune your parameter thresholds. Your customer base may be comprised of different tiers, such as personal accounts versus commercial accounts; or small business, medium sized, and large accounts. You can use the simulation tool to fine tune the scenario thresholds for each of the different account tiers.

Once you are satisfied that the SAS Visual Scenario Designer scenarios are properly tested and tuned, the latest version of your SAS Visual Scenario Designer deployment can be seamlessly incorporated into the SAS Anti-Money Laundering alert generation process. The alert generation process executes the windows and scenarios that have been enabled within the deployment on the high performance in-memory LASR server. The results are ported into the necessary format and appended into the SAS Anti-Money Laundering alert list. The *SAS Anti-Money Laundering 6.3: Installation, Configuration and Administration Guide, Section Edition* provides detailed information about this process.

The SAS Anti-Money Laundering web-based user interface supports the management, investigation, and reporting needs of anti-money laundering analysts and investigators. The solution includes a banking specific data model; provides suspicious activity monitoring and reporting; investigation and alert management; customer due diligence including KYC risk scoring and classification; watch-list matching; and an integrated case management. The new Entity Triage window facilitates triaging at the Customer level and provides all the data necessary to triage a decision about a customer through this interface, rather than having to delve into multiple screens.

## CONCLUSION

SAS Visual Scenario Designer can now be integrated with the SAS Anti-Money Laundering solution. As a result, scenarios can be created, tested, tuned, and simulated using a point and click interface. An audit trail is logged and available for auditors. Once deployed, the SAS Visual Scenario Designer scenarios execute on a high speed in-memory LASR server in conjunction with the traditional anti-money laundering scenarios. The alerts generated can be viewed and investigated through the SAS Anti-Money Laundering GUI. The automated business process workflow ensures that the correct progression of tasks is adhered to while providing a detailed audit trail of activities. Cases can be created and managed through the Case Management tool. Using the SAS Visual Scenario Designer in tandem with the SAS Anti-Money Laundering solution offers an effective mechanism to detection anomalous and nefarious behavior; meeting compliance requirements; and safeguarding your institution's reputation while lowering development and maintenance costs.

## REFERENCES


Author name: Rexrode, Christina. December 10, 2014. "Citi Sees Legal Charges", The Wall Street Journal.

Author name: Colchester, Max. November 26, 2015. MarketWatch. Available at http://www.marketwatch.com/story/barclays-fined-for-anti-money-laundering-failings-2015-11-26

Contract: Hudak, SteveFinCEN, January 27 2015. Available at: https://www.fincen.gov/news_room/nr/pdf/20150127.pdf

SAS Visual Scenario Designer 6.3: User's Guide, Second Edition, Cary NC: SAS Institute Inc. Available at http://supportprod.unx.sas.com/documentation/solutions/vsdesigner/6.3/vsdug.pdf

SAS Visual Scenario Designer 6.3: Administrator's Guide, Second Edition, Cary NC: SAS Institute Inc. Available at: http://supportprod.unx.sas.com/documentation/solutions/vsdesigner/6.3/vsdag.pdf

SAS Anti-Money Laundering 6.3: Installation, Configuration and Administration Guide, Second Edition, Cary NC: SAS Institute Inc. Available at http://rnd.sas.com/sespu/aml/Shared%20RD%20Documents/amlsag-6.3m1.pdf

SAS Anti-Money Laundering 6.3: Scenario Administration User's Guide, Second Edition, Cary NC: SAS Institute Inc. Available at http://rnd.sas.com/sespu/aml/Shared%20RD%20Documents/amlscug-6.3m1.pdf

SAS Anti-Money Laundering 6.3: User's Guide, Second Edition, Cary NC: SAS Institute Inc. Available at http://rnd.sas.com/sespu/aml/Shared%20RD%20Documents/amlinvug-6.3m1.pdf


## ACKNOWLEDGMENTS


- The SAS Anti-Money Laundering – Fraud and Financial Crimes R&D Team
- The SAS Visual Scenario Designer – Fraud and Financial Crimes R&D Team


## RECOMMENDED READING

- SAS Visual Scenario Designer 6.3: User's Guide, Second Edition
- SAS Visual Scenario Designer 6.3: Administrator's Guide, Second Edition

- SAS Anti-Money Laundering 6.3: Installation, Configuration and Administration Guide, Second Edition
- SAS Anti-Money Laundering 6.3: Scenario Administration User's Guide, Second Edition
- SAS Anti-Money Laundering 6.3: User's Guide, Second Edition

## CONTACT INFORMATION

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

Renee Elizabeth Palmer
Principal Solutions Architect
Security Intelligence Practice
SAS Institute Inc.
SAS World Headquarters
919-531-1929
Renee.Palmer@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.

SAS5801-2016

# Minimizing Fraud Risk through Dynamic Entity Resolution and Network Analysis

Danielle Davis, Stephen Boyd, and Ray Ong, SAS Institute Inc., Cary, NC

## ABSTRACT

Every day, businesses have to remain vigilant of fraudulent activity, which threatens customers, partners, employees, and financials. Normally, networks of people or groups perpetrate deviant activity. Finding these connections is now made easier for analysts with SAS® Visual Investigator (SAS® VI), an upcoming SAS® solution that ultimately minimizes the loss of money and preserves mutual trust among its shareholders. SAS Visual Investigator takes advantage of the capabilities of our new SAS® Viya™ server. Investigators can efficiently investigate suspicious cases across business lines, which has traditionally been difficult. However, the time required to collect, process and identify emerging fraud and compliance issues has been costly. Making proactive analysis accessible to analysts is now more important than ever. SAS Visual Investigator was designed with this goal in mind and a key component is the visual social network view. This paper discusses how the network analysis view of SAS Visual Investigator, with all its dynamic visual capabilities, can make the investigative process more informative and efficient.

## INTRODUCTION

Today's fraud is like a moving target. It purposely flies under the radar and follows the path of least resistance. With companies automating more of their processes, the window to detect fraud is also shrinking. Traditional techniques often fail to identify fraudulent behavior. According to Gartner Research, basic monitoring for deviation will select the sole fraudsters but it falls short when looking at fraud rings. Also can be prone to a lot of false positives, which is not good for customer service.

SAS® VI is a solution to help your company detect and manage fraud. The social network view of SAS® VI can provide useful insight into large data sets along network, spatial, and time dimensions based on the interconnectedness of the subjects being analyzed. This social network analysis (SNA) can be a key component when looking for fraudulent activity.

### FRAUD DETECTION

Why is fraud hard to detect? First, it is an uncommon pattern. It will always take the path of least resistance and the whole goal is to fly under the radar, to blend in. Therefore, there is rarely frequently occurring patterns where classical data mining techniques can be used. Second, today's fraud is typically very carefully organized crime. They do not operate independently, they are big organizations with multiple players and roles being played by different individuals. However, fraudsters try to blend into the environment and not behave different from others in order not to get noticed and to remain covered by non-fraudsters. Third, techniques and tricks fraudsters evolve over time. It looks to be an endless game of cat and mouse played between fraudsters and fraud fighters. Lastly, instead of setting up a massive, one-time strike, they can do a lot of smaller activities under the radar for the same monetary result. So how can we battle these issues? SAS® VI social network view can help ease the pain of detecting fraud.

## SOCIAL NETWORK VIEWER

A social network view of your data is designed to analyze the data to identify relationships, and mining it for new information (such as the quality or effectiveness of a relationship). It identifies patterns of behavior that only appear suspicious when viewed across related accounts or attributes of an account. You can examine social structure and interdependencies (or work patterns) of individuals or organizations.

Let's breakdown the viewer into components that can assist in revealing fraud.
First, let's look at the viewer from a high level:

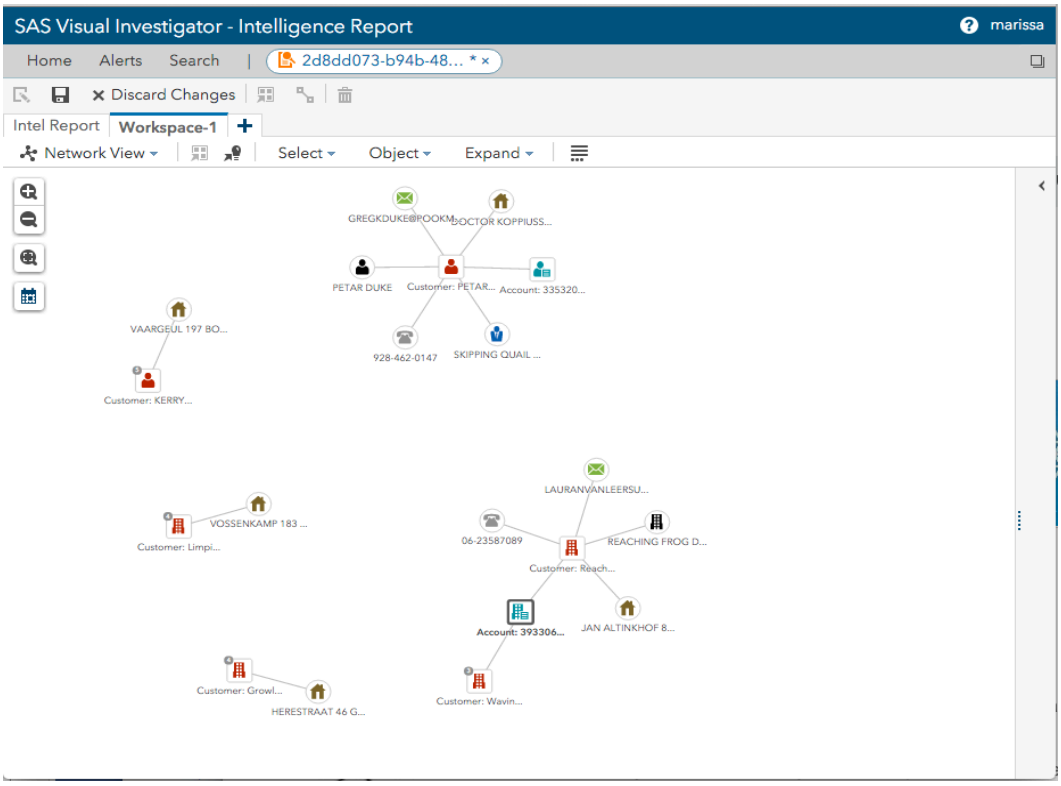



**Figure 1 - SAS VI - Social Network View**

## SEARCH AND EXPANSION

In order to detect the unusual patterns of fraud, underlying fraud rings and fraud that crosses companies and/or product lines, a user must be able link nodes together. In other words, an analyst needs to see the relationships between information/knowledge entities. Since the network viewer is only as good as its data, SAS® Visual Investigator's network viewer is powered by data that has been processed through entity resolution and then stored and indexed in Elasticsearch. This provides the user the ability to have efficient searches for resolved document information that would have been otherwise overlooked.

The Entity Resolution Service enables SAS® Visual Investigator to identify unique entities across various source documents, which can have different data representing the same entity. Particularly in fraud detection scenarios, a Person entity, for example, can represent themselves with different first names, different last names, different phone numbers, different addresses, and different SSNs. Entity resolution seeks to find matches across a number of different compounds, each compound being defined as a set of columns (elements) that should determine uniqueness.

Elasticsearch storage provides quick retrieval time and allows the user to easily add new nodes to their existing network and discover relationships that were previously undetected. This also allows the analyst to know if a node has additional relationships that can be expanded on the network. This expansion can expand one or 2 levels as well as providing a complete community expansion.

Below is an example of the search feature of SAS® Visual Investigator in order to populate the network viewer with relational data. Once a user has a list of possibilities in his search, he/she can add these entities to a workspace. This is where the user can start to explore the relationships.

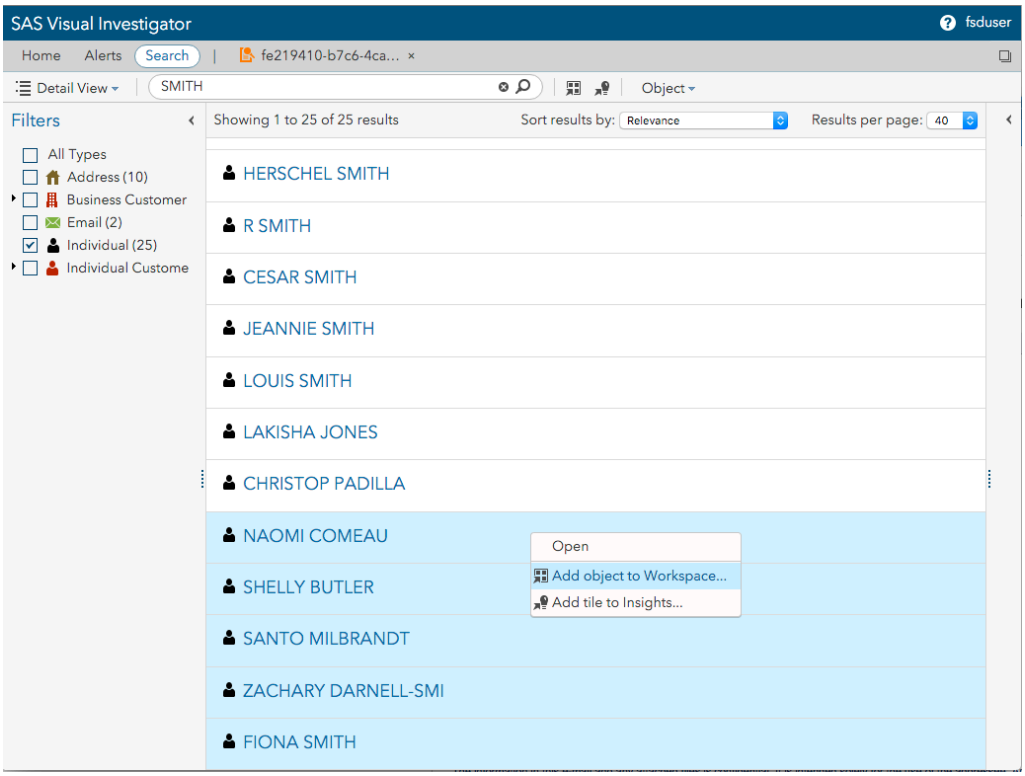



**Figure 2. Search and Entity Resolution Built around the Network Viewer**

Once the entities and documents are loaded in a workspace, the user can visually tell which nodes have further relationships that can be explored. These nodes contain a degree in the upper left corner to indicate the number of additional direct relationship they have. These nodes can be expanded in the network to reveal the additional information.

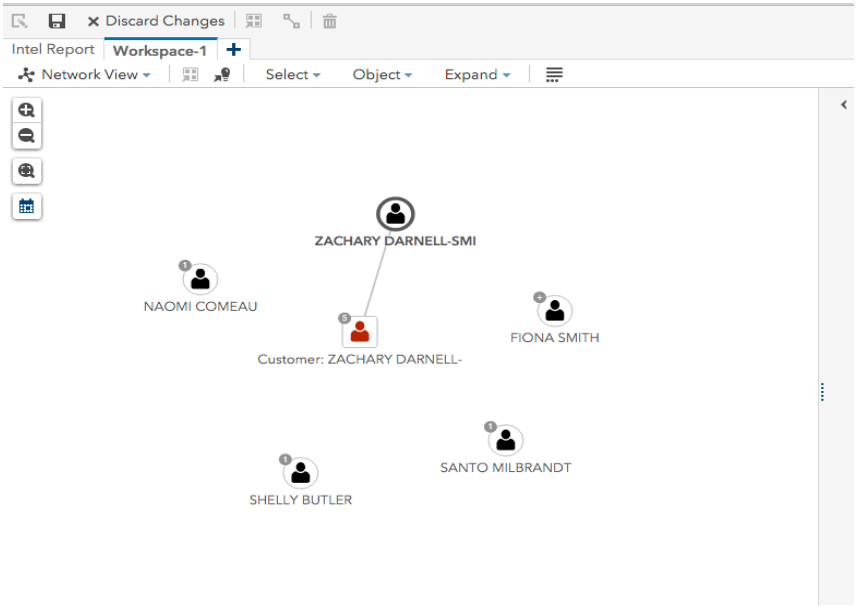



**Figure 3. Five Nodes in the Network Have More Direct Relationships**

This functionality allows a user to begin to understand the relationships that might otherwise go unnoticed. SAS® Visual Investigator adds another level of analytical power by using SAS® Viya™ and allowing a user to expose the entire community of a particular node. In other words, using analytical techniques to expose nodes that are highly interconnected to the node selected.

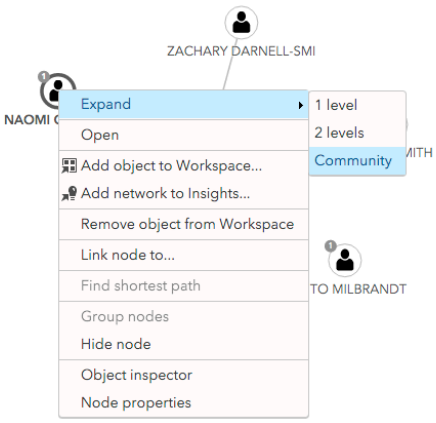



**Figure 4. Expanding a Node**

## NODE AND LINK PROPERTIES AND ANNOTATION

Part of what makes a network viewer essential to fraud detection is its ability to help the analyst tell/track the fraud pattern in the network. This requires customization of the network nodes and links.

The network view in SAS® Visual Investigator allows user to change many different properties for a node including its size, color, shape, and icon.

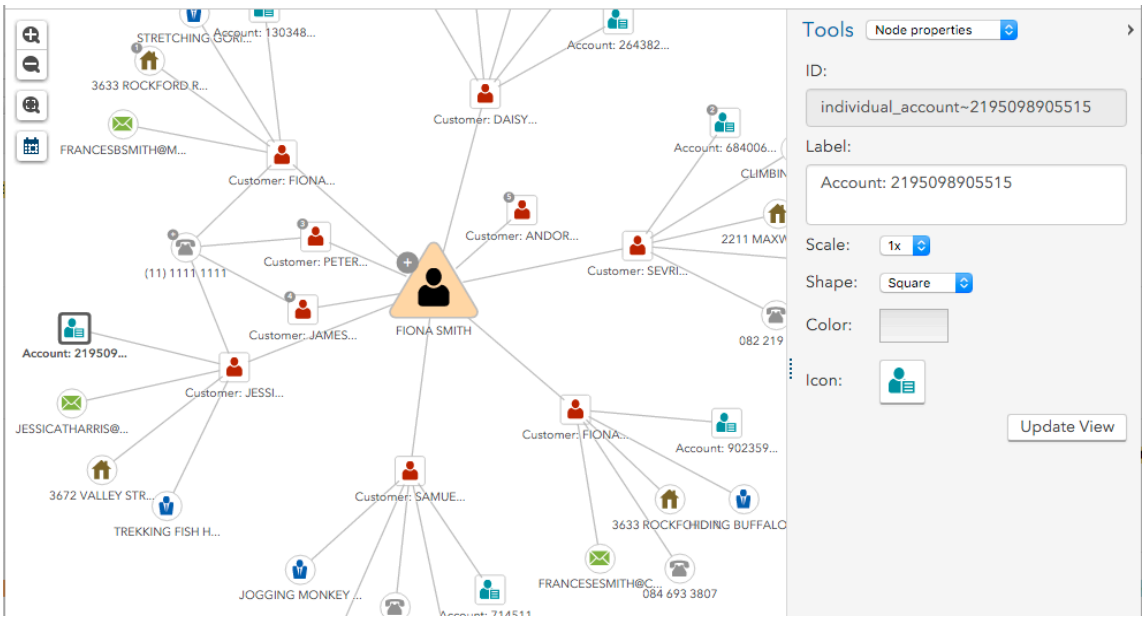



**Figure 5. Node Properties**

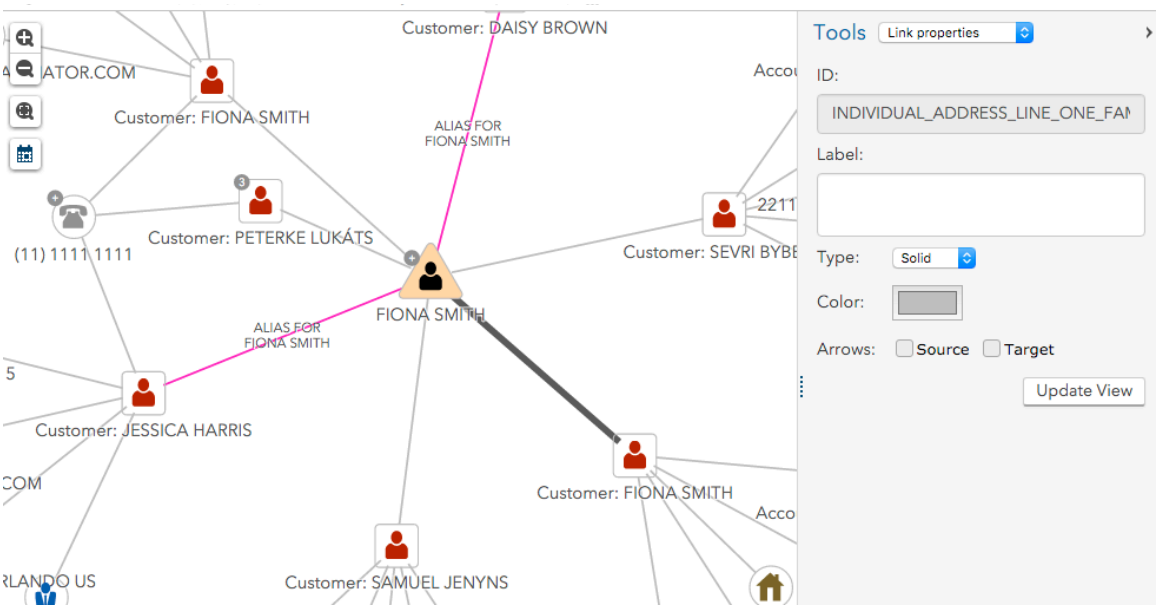



**Figure 6. Link Properties**

Node annotation is another feature that can assist an analyst in a number of different ways. These annotations are added based on the value of a specific attribute on the entity or document. This can be beneficial with the following:

- To help further document the investigation of the potential fraudulent activity.
- To alert the analyst that the entity has exceeded a threshold.
- To further define a node in more specific terms. For example, if there is a claim the annotations could inform the analyst of the type of claim, claim status, and insurer for the claim.

The annotations can be placed on 8 different locations of the nodes and can be in the form of text or icons.


**Figure 7. Icon Decoration**

## GROUPING NODES

Another key feature in helping the analyst understand the story of the fraudulent activities is the ability to group nodes. The grouping functionality not only allows you to “clean up” you network without discarding important information it allows to view the contents of these nodes through our “object inspector”, which shows detailed information about a node.

If you look back at the links that are referenced in Figure 6, we see that they reference two nodes that are aliases for another person. A user can group all of those nodes together so that it helps minimize the complication of the network.

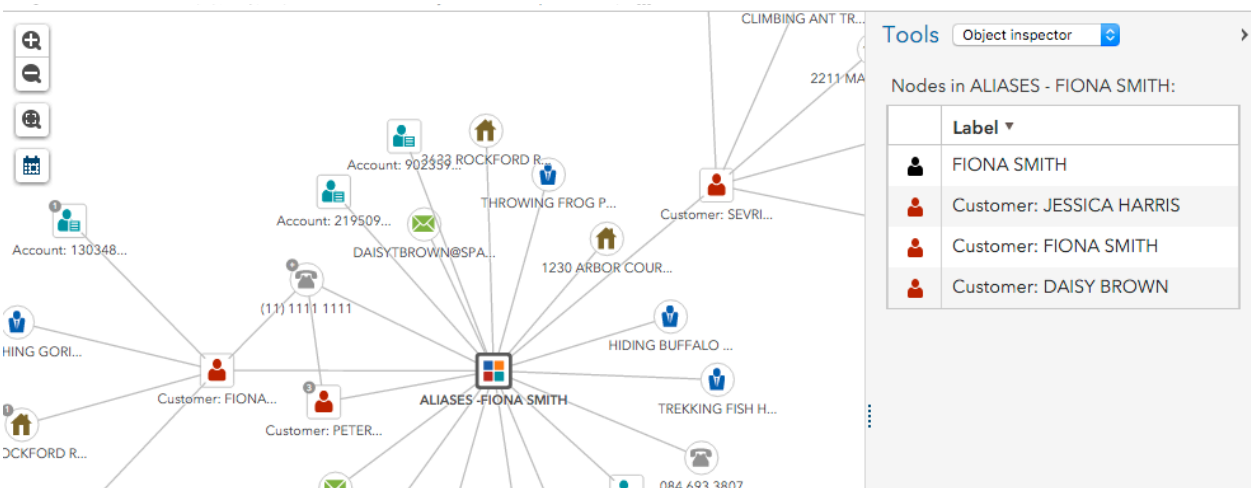



**Figure 8. Grouping Nodes**

## NODE SELECTION

When networks become large, it is very cumbersome and tedious to find specific nodes. SAS® Visual Investigator provides a tool to easily find nodes that are of the same type and an option to search for a string in any of the name labels.

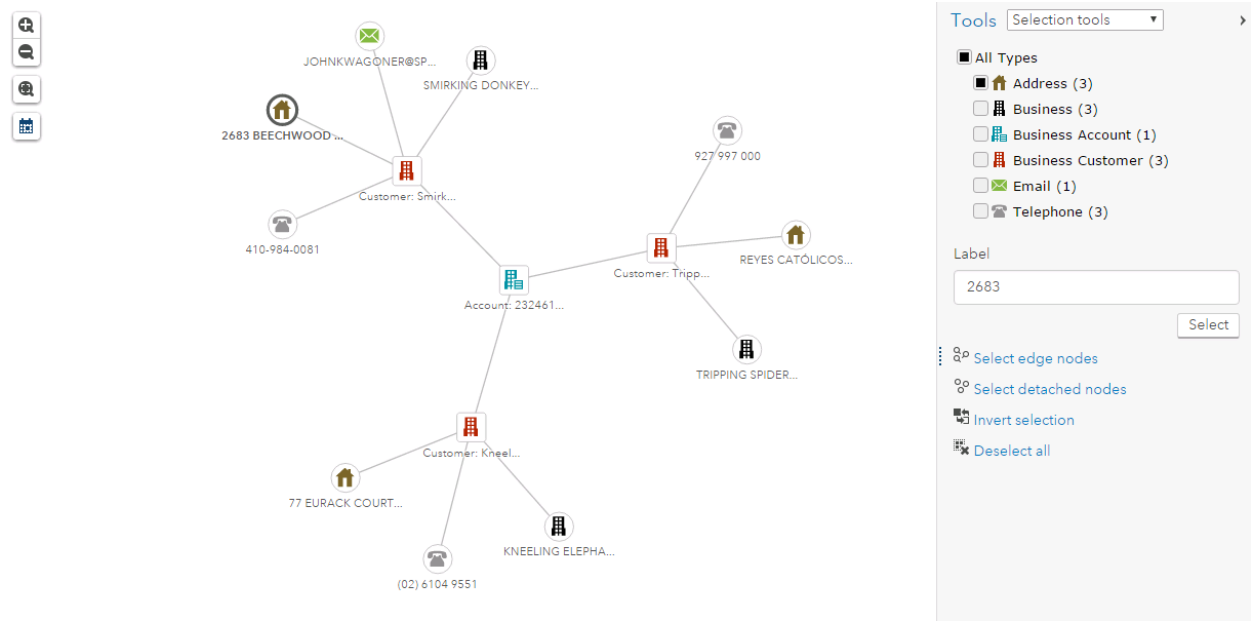



**Figure 9. Node Selection**

## NETWORK LAYOUT

Some fraud rings can be quite large and this can require your social network to get quite large and adding new information to an existing network can become tedious when it comes to laying out all the nodes. SAS® Visual Investigator's network view allows the user to layout and re-adjust their network in a number of different ways.

1. Initial layout tries to ensure that nodes are not overlapping and minimizes the amount of link crossing. This applies to new nodes coming into the network as well as new nodes coming in from an expansion. The entire network might need to be adjusted depending on how many nodes are being added.
2. If a user has spent quite a bit of time adjusting individual nodes, they can also choose to only adjust the nodes that have not been manually moved.
3. A user also has the ability to choose to adjust only the nodes they have selected in the network.
4. Advanced features are also available for over all fine-tuning of the network.

The network viewer can also handle networks with thousands of nodes by reducing the details so that the analyst can see the overall view of the network as well zoom into the details.

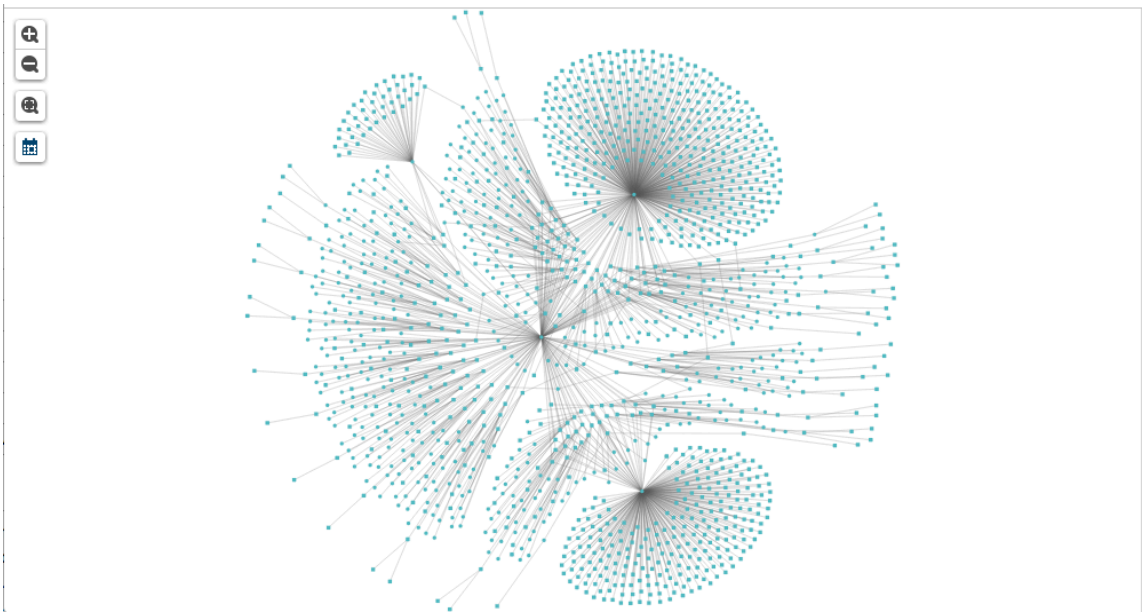


**Figure 10. Large Networks**

## NETWORK ANALYTICS

Social Network Analysis can also involve some statistical analytics that can assist with understand the important players in a network. Taking advantage of SAS® Viya, SAS® Visual Investigator's network viewer provides several key network analytics that can help answer questions:

- How highly connected is an entity within a network?
- What is an entity's overall importance in a network?
- How central is an entity within a network?
- How does information flow within a network?

## TIMELINE

Another key feature to network analysis is to understand when particular entities started their role in the network. The ability to look at patterns over time. The network viewer contains a time slider component that can provide a simple view of communications that happened within a given time-frame and a continuous temporal view that includes the history before the time-frame defined. Let's review both options.

The diagram below depicts the time slider in a "no-history" mode:

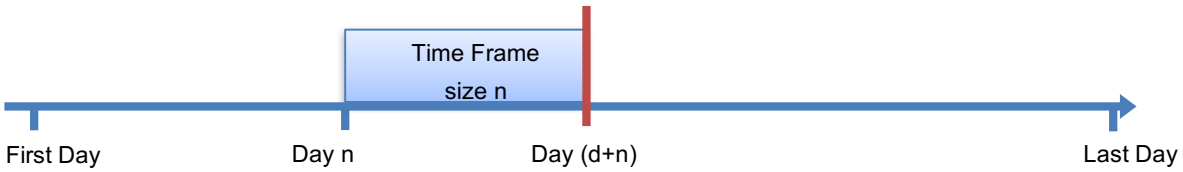



**Figure 11. Time Slider in "no history" Mode**

In this example, day d is the first day that the visualization is showing and the current time frame is [d, d+n]. Only communications inside the current time frame are calculated and displayed, and only communications within that time-frame are considered active.

The following figure shows the time-slider in the "no-history" mode:

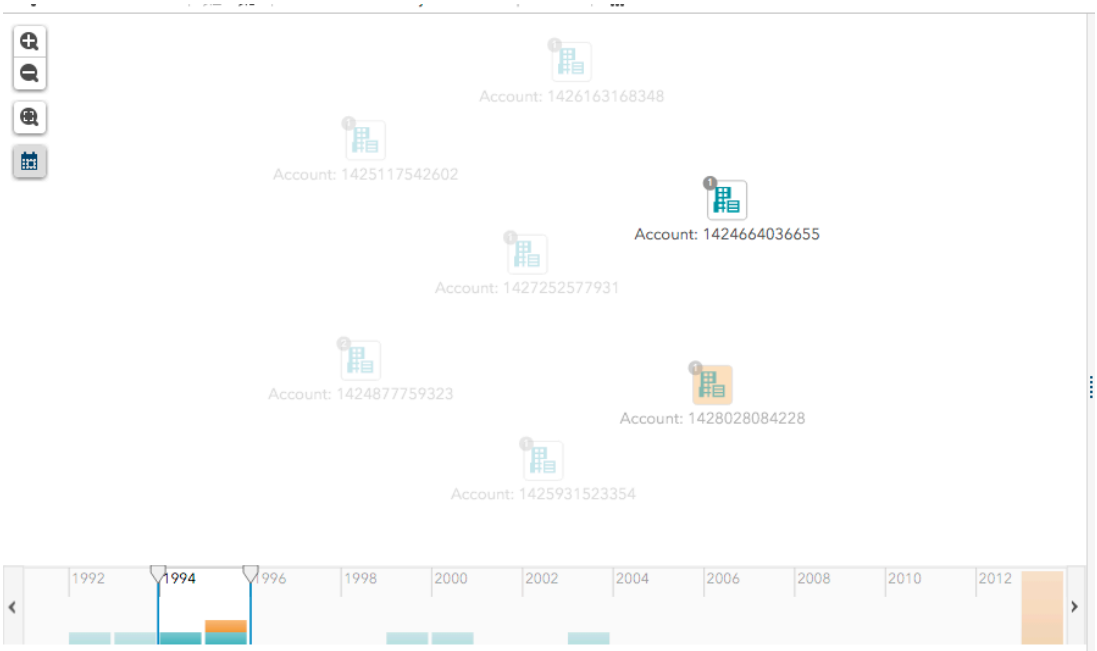



**Figure 12. Time Slider with "no-history" Mode**

After calculating the full range of dates, the timeline will determine what is the best interval to "bucket" date values in. Above you see that the range of data was over 10 years so that the intervals are yearly buckets. Notice that the nodes that are grayed out have start dates outside of our selected view, whether they started before the date range or after. The nodes that are grayed out but also have a muted orange background are nodes that had end dates within our range. In other words, they had an end date in 1995.

The next diagram shows the time slider in the “with history” mode:

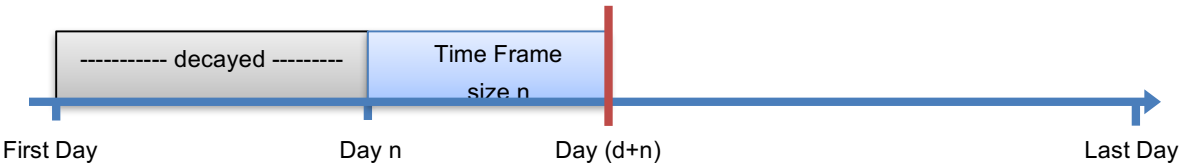



**Figure 13. Time Slider in "with history" Mode**

This time frame window allows users to foresee the activities happening inside the time frame after the current day. Thus, day d is the current day that the visualization is showing and the current time frame is [d, d+n]. All communications through day (d+n) are calculated and displayed, and if a communication takes place before or on day d, it is active. Below is an example of the time-slider using the “with history” feature:

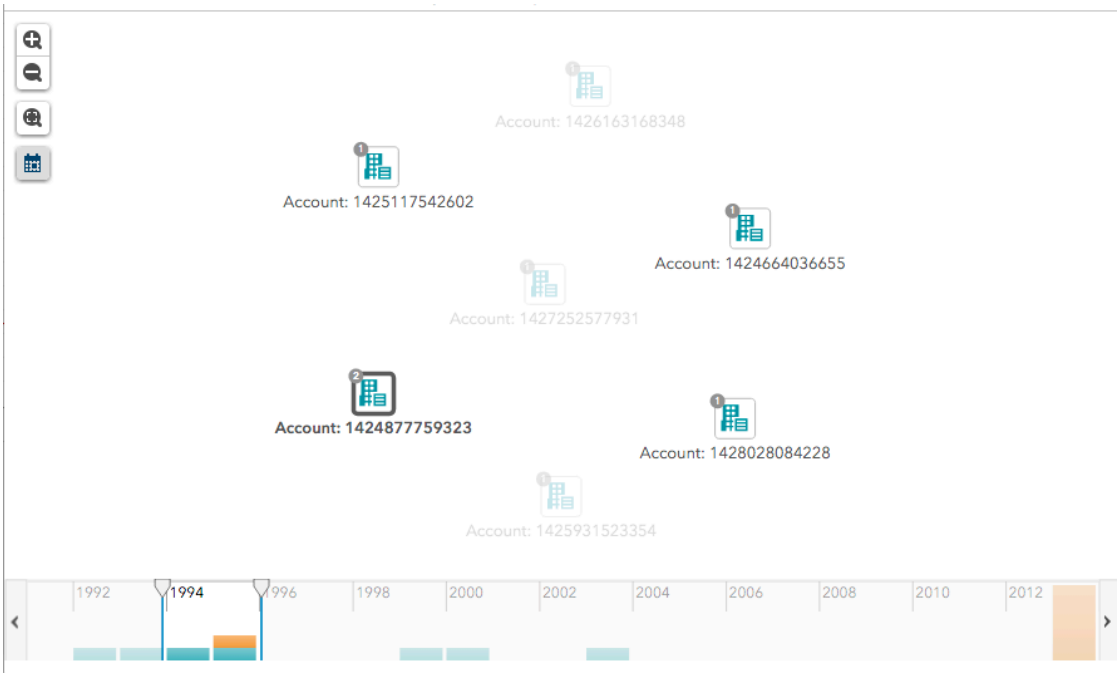



**Figure 14 . Time Slider Show "with history" Feature**

Notice that the nodes that are grayed out have not come into play yet. The have start dates after 1995. However, all other nodes have start dates within or before the range of time we have in view. None of the nodes are grayed-out and orange since the left-handed slide has not passed an orange bar in the timeline. Hence, the end date is not in the past yet.

The timeline also allows the user to zoom in on a bar and see the actual dates that fell into that bucket and those nodes the network will reflect that zoomed in view.

## CONCLUSION

Social network analysis is an important part of a layered prevention approach to fraud and compliance discovery. A complete fraud solution involves incorporating fraud and compliance modeling to prevent fraudulent activity from surfacing in line, and social network analysis to discover new emerging patterns or changes in known fraud and compliance patterns. This combination provides a solution that can quickly update the fraud and compliance alert system with higher quality rule sets for triggering potential fraud and decreasing false positives or customer friction.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Danielle Davis
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Danielle.Davis@sas.com
http://www.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.

Paper SAS6800-2016

# Alerts Don't Launder Money (or Finance Terrorism) – People Do!

Kathy Hart, Malcolm Alexander, SAS® Institute Inc.

## ABSTRACT

For far too long, anti-money laundering and terrorist financing solutions have forced analysts to wade through oceans of transactions and alerted work items (alerts). Alert-centered analysis is both ineffective and costly. The goal of an anti-money laundering program is to reduce the risk for your financial institution, and to do this most effectively, you must start with analysis at the customer level, rather than simply troll through volumes of alerts and transactions. This paper discusses how a customer-centric approach leads to increased analyst efficiency and streamlined investigations. Rather than starting with alerts and transactions, starting with a customer-centric view allows your analysts to rapidly triage suspicious activities, prioritize work, and quickly move to investigating the highest risk customer activities.

## INTRODUCTION

Money laundering is the process of making illegally gained proceeds appear legal. To protect the financial system from "the abuses of financial crime, including terrorist financing, money laundering and other illicit activity", the United States established the Bank Secrecy Act (BSA) in1970. The European Union established directives on anti-money laundering in 1991. These statutes and the follow-on legislation require banks and other financial institutions to report suspicious activity to the appropriate authorities.

These legislations require financial institutions to implement an Anti-Money Laundering Compliance Program to monitor for suspicious activity and then file Suspicious Activity Reports (SAR) in the U.S. Other countries have their own set of reports that they must file. Financial institutions use software to automate monitoring and to generate alerts when suspicious activity occurs. Analysts investigate the alerts and decide whether to close the alert or to create a case for an investigator to review and disposition. The investigator gathers additional information before deciding whether to close the case or file a SAR.

SAS® released its first anti-money laundering solution, SAS Anti-Money Laundering, more than 10 years ago. SAS Anti-Money Laundering uses scenarios to generate alerts that are then reviewed by the analysts.

The solution has evolved with the ever-changing landscape of the global market. Originally, Anti-Money Laundering Compliance programs focused on dispositioning alerts, which represent granular behaviors, such as out-of-footprint ATM activity. Overtime, money-laundering techniques have become more complex and the regulatory landscape has changed. Creating more scenarios for a broader monitoring program are now the norm. These changes result in an all-time high for alert generation.

As a result of these changes, a more holistic approach to investigating money laundering is required. This paper presents some major changes and improvements to SAS Anti-Money Laundering that make detection more meaningful, triage and investigation more efficient, and greatly simplify the day-to-day management of Anti-Money Laundering Compliance programs in financial institutions.

## TRADITIONAL ALERT TRIAGE AND CASE INVESTIGATION

### STANDARD TRIAGE AND INVESTIGATION PROCESS

The goal of this effort is to come to a decision as to whether to report a customer or external party that does business with the bank (both referred to as an entity), as suspicious.

Figure 1 shows the general process:

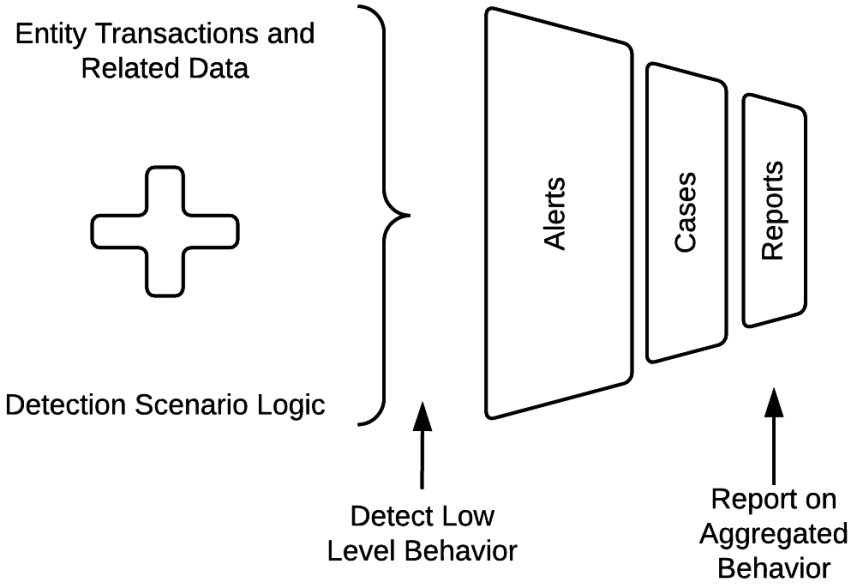



**Figure 1. Typical Transaction Monitoring Process**

In a typical process, the software uses entity and scenario information to generate alerts that represent suspicious behavior at a granular level. Table 1 contains examples of granular alerts generated by Anti-Money Laundering scenarios.

**Table 1: Example Alerts**

| Alert Level | Alert Category | Alert Type | Alert Description |
|---|---|---|---|
| Account | Cash | Structured Withdrawals | An account has multiple large withdrawals over a short period. |
| Customer | Watch list | Politically Exposed Person | A customer is designated as a politically exposed person (PEP) on a watch list |
| Household | Cash | Large Cash Deposits | Customers in a household have a daily total amount of cash deposits that exceed a threshold |
| Associates | Other | Balance Inquiry | An associate performs a large amount of balance inquiries |
| External | Wires | Multiple Internal Beneficiaries | A single external remitter wires money to multiple internal beneficiaries. |

Analysts then use an application to review, or triage these low-level alerts, working to determine whether the detected transactional behavior is worthy of a deeper investigation. The alerts generated are associated with customers, accounts, households, associates, or external parties. If it is determined that the alert requires additional review, the analyst creates a case.

Cases bridge the divide between low-level behavior detected by automated systems, and the regulatory report that is filed on an individual or organization. Investigators trained in the research and analysis required to detect suspicious behavior perform the more complicated work required for cases. Where

analysts traditionally look at an alert and its associated transactions in relative isolation, investigators pull together the whole picture of individual or group activity related to money laundering or terror financing.

Herein lies the major challenge. Given the growing sophistication of criminals, we detect granular behavior, but ultimately report on the aggregated behavior of people, groups of people, or organizations.

### CHALLENGES WITH THE TRADITIONAL WAY

The mismatch between the ways that we detect and how we report has important consequences. On the business side, a financial institution risks under-reporting because at the beginning of the process, triage, analysts miss broader patterns of behavior that span alert types. Even with additional training and best practices, analysts must actively search out other alerts related to the entity under review. This is a manual process and thus prone to errors. Overly conservative analysts create extra work for investigators, pushing alerts on legitimate transactions (also known as false positives) to cases. In this process, investigators are the first to take a holistic look at the entities, which increases their workload.

Since dispositioning each alert individually is time consuming, tuning the scenarios to reduce the number of false positive alerts is very important. However, tuning scenarios to reduce false positive alerts can actually suppress the generation of alerts on truly suspicious activities.

Financial institutions deal with the workload issues in a variety of ways – none of them optimal. Some organizations organize by alert “type”, for example, routing all wire alerts to a special sub-team. This, and various other ways of organizing around alerts has the side effect of causing workload balancing headaches – and is still subject to the risk of missing aggregate behavior. Others create “super-alerts” in an effort to bend an alert-oriented system to the need for broader investigation. This causes some loss of visibility of the important lower-level behavior, and its efficacy is subject to the degree to which you can automatically roll up behavior to a meaningful super-alert.

On the technical side, systems built end-to-end around granular behavior tend to be cumbersome and difficult to use when trying to synthesize information for the purposes of investigation. Built around the alert, facilities for quickly navigating to other entities, switching back-and-forth to gain a broader understanding of behavior, or viewing both detailed and broader behavior can be unwieldy. Visualizations of alert-level information are not as useful as they might initially appear.

In summary, the implications of working with alerts, but reporting on entities, are far-reaching and touch many areas of a BSA program. The remainder of this paper draws on these organizational and technical lessons to propose a new approach, which is being implemented for SAS Anti-Money Laundering 7.1.

## A BETTER APPROACH

A change in philosophy to make people the focus, from the beginning, resolves many of the challenges. Figure 2 shows the new philosophy.

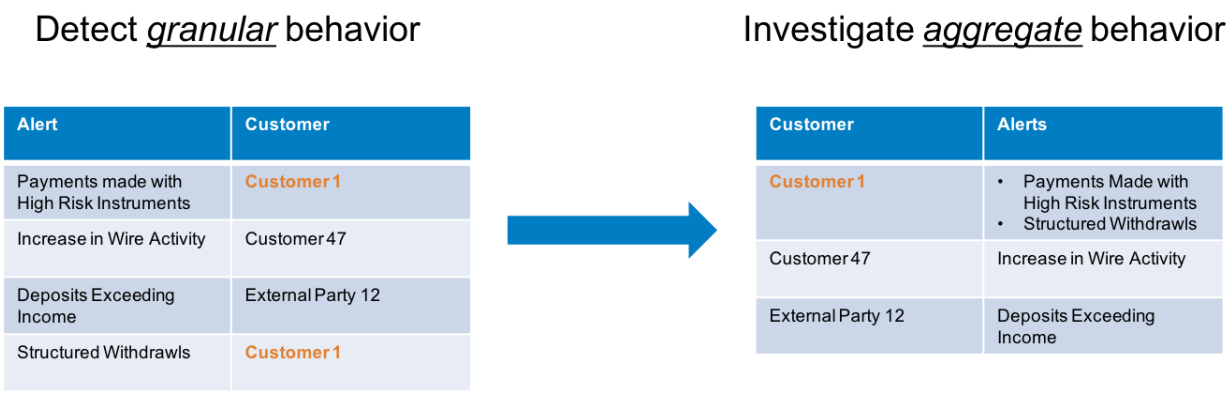



**Figure 2: Entity Triage**

The goal here is to retain the benefits of detecting low-level behaviors, but gain the ability to look holistically at an entity and quickly come to a decision as to whether their overall behavior is suspicious. Beyond just new reports and aggregations, moving to this philosophy has implications for the processes that generate alerts, and opens up new opportunities to provide much more streamlined and efficient user interfaces. This makes analysts and investigators more productive.

## ALERT GENERATION

SAS Anti-Money Laundering is known for its white-box approach to transaction monitoring; financial institutions globally use customizable out-of-the-box scenarios, or develop their own scenarios to detect granular behavior. The engine of this detection process is the Alert Generation Process, or AGP.

The AGP examines entity transactions and related data, running scenarios to generate alerts. Transitioning to an entity-centric view affects numerous parts of the AGP.

### Aggregation

After the alert generation is complete, the alerts are aggregated to a customer, associate, bank, or external party for further investigation and disposition. Aggregation rolls up account alerts to the primary owner of the account and household alerts to the head of household. After aggregation, triage of all alerts is possible by reviewing people or organizations. Table 2 shows the rules for aggregating alerts to entities.

**Table 2: Alert to Entity Mapping**

| Alert Type | Entity to which alerts Aggregated |
|---|---|
| Account | Customer |
| Customer | -- |
| Household | Head of Household Customer |
| External Party Account | External Party |
| Bank | Bank |
| Associate | Associate |

### Replication of Data

Investigating an alert includes investigating transactions associated with the alert. This includes not only the transactions that cause the alert but also other related transactions. Replication is the process of gathering the related transactions for investigating and archiving. Changing investigations from alerts to people means investigating transactions related to the suspicious people. For account alerts, replication gathers all transactions associated with the primary account owner and not just transactions associated with the account. For household alerts, replication associates transactions from all accounts in the household to the customer defined as the head of the household.

### Alert Routing

Since investigators no longer receive individual alerts, routing of alerts to investigators based on alert type rules is no longer necessary. Instead of matching alerts with specialists, automatic routing is available for workload balancing. A round robin assignment method determines the assignment of customers and other entities with alerts to investigators. The assignment maintains ownership of existing entities in triage and assigns new entities to investigators in alphabetical order.

Routing of aggregated information allows more well-rounded analysts versed in a variety of suspicious behavior to produce effective and productive cases.

### AGP Summary

Aggregation, replication, and routing are challenging for customers or consultants to bolt-on; the system has to be built with this in mind, which is the major effort related to the AGP in SAS Anti-Money Laundering 7.1.

### USER INTERFACE

SAS Anti-Money Laundering 7.1 extends the entity triage component of SAS Anti-Money Laundering 6.3m1 to provide a completely new interface to take advantage of the aggregations provided by the AGP. A few key principles drive the design of the new user interface:

- Leverage the aggregated information to display holistic views of entities.
- Display the most critical information prominently, focusing on those items key to making decisions.
- Simplify and flatten navigation, allowing a variety of entities and other objects to be viewed at the same time

### Holistic Views

Rather than reviewing individual alerts, the Triage component for example, provides aggregated views. Figure 3 shows the entity triage list.

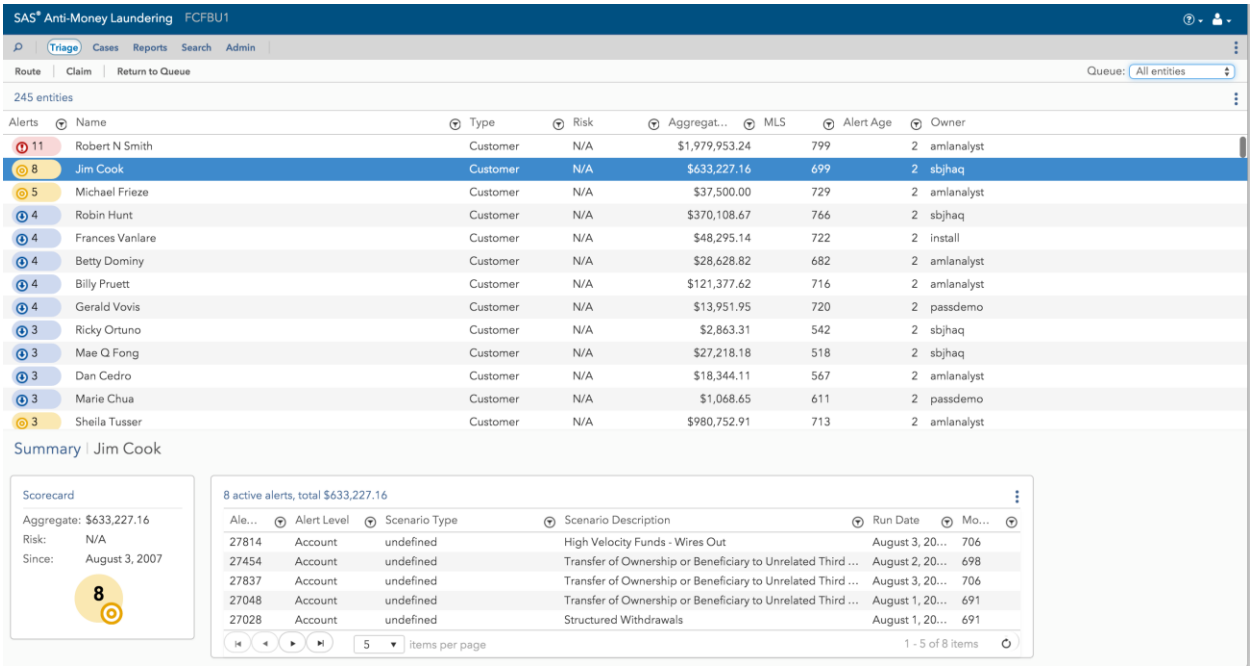



**Figure 3: Entity Triage**

Rather than prioritizing by alert type, new metrics are available to better quantify risk to your institution across all entities, whether they are internal customers, external parties, or any of the other types noted in Table 2. The user can customize the prioritization through filtering and sorting of the metrics. The application retains these settings across user log-ins, and provides an option to reset to the defaults. Selecting a row in the detail screen shows a summary, at the bottom of the screen, of the underlying alerts. Double-clicking on a row displays the detail page

### Focus on Decision Making

Each page in the application presents the most useful information for making a decision at a particular time. The customer detail screen shown in Figure 4, for example, displays all the open customer alerts, linked to the transactions table at the top.

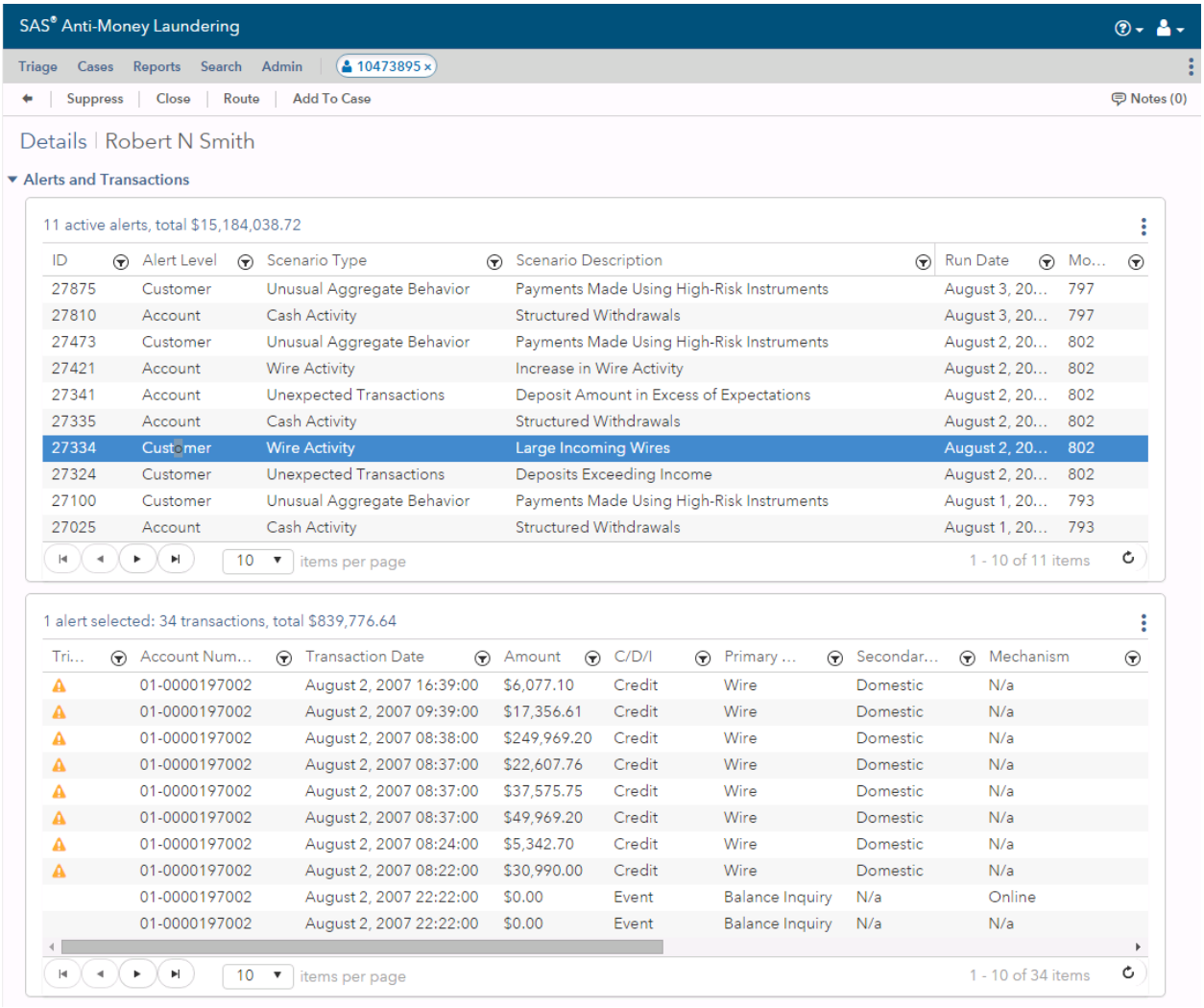



**Figure 4: Customer Alerts and Transactions**

Selecting an alert dynamically changes the transaction view to filter to the triggering and related transactions for that alert. De-selecting an alert brings back the view of all the transactions available. With these linked tables, users can quickly find transactions of interest. Demographic information, typically found at the top of the page in most applications, is less important to the actual decision and thus placed farther down the page.

### Navigation

Systems designed around alerts tend to use a strict page hierarchy that encourages drill-down, and are not optimized for reviewing a variety of different entities at once. SAS Anti-Money Laundering 7.1 flattens the navigation to allow users to see and access a variety of information very quickly. The screenshot in Figure 5 highlights some of these features:

- Main areas of the application are always available with one click (Triage, Case, Report, and so on.)
- A variety of items can be open at the same time, and are cached to avoid repeated database calls and performance challenges. For example, a Customer, a Transaction, an External Party, and an Account detail page can all be open and users can switch back-and-forth easily – without having to leave the case or triage item that they are working on.
- Any actions that the user can take are always available, regardless of where you are on the page. Pages scroll vertically to provide quick access to lots of information. The menu bar with buttons like “Suppress” and “Add to Case” are “sticky” – that is, they are always available at the top of the screen.

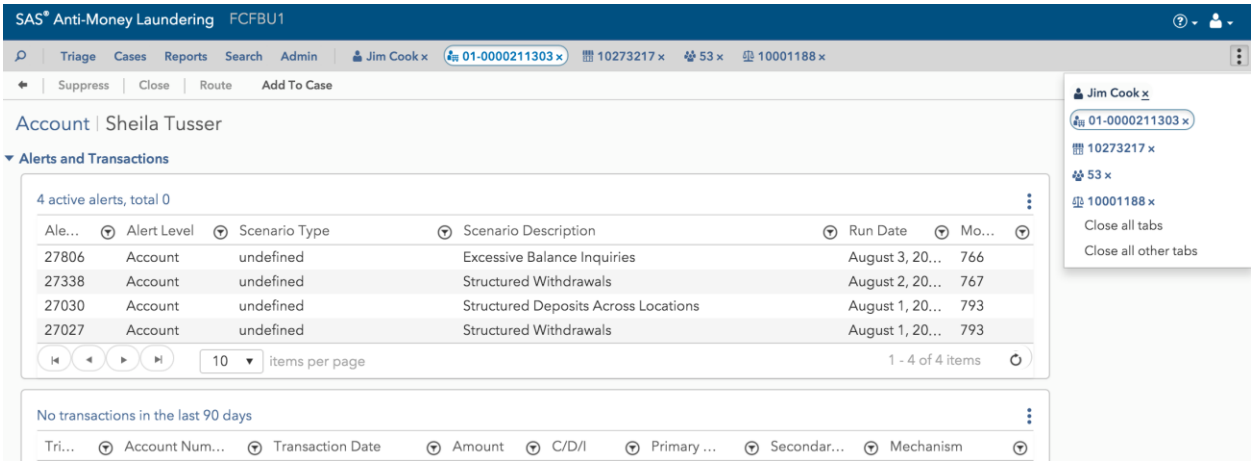



**Figure 5: Navigation Methods**

## CONCLUSION

The compliance space has transitioned from an era of process adherence to an era that places great demands on the ability of organizations to identify complex activities by possible multiple actors. Working at the level of the individual alert is no longer viable, and SAS Anti-Money Laundering 7.1 makes BSA organizations more effective.

## ACKNOWLEDGMENTS

Thanks to Brian Ferro and Dan Tamburro.

## CONTACT INFORMATION

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

Kathy Hart
Kathy.Hart@sas.com

Malcolm Alexander
Malcolm.Alexander@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.

**Paper SAS524-2017**

# Adding a Workflow to Your Analytics with SAS® Visual Investigator

Gordon Robinson and Ryan Schmiedl, SAS Institute Inc.

## ABSTRACT

Monitoring server events to proactively identify future outages. Looking at financial transactions to check for money laundering. Analyzing insurance claims to detect fraud. These are all examples of the many applications that can use the power of SAS® Analytics to identify threats to a business. Using SAS® Visual Investigator, users can now add a workflow to control how these threats are managed. Using the administrative tools provided, users can visually design the workflow that the threat would be routed through. In this way, the administrator can control the tasks within the workflow, as well as which users or groups those tasks are assigned to. This paper walks through an example of using the administrative tools of SAS Visual Investigator to create a ticketing system in response to threats to a business. It shows how SAS Visual Investigator can easily be adapted to meet the changing nature of the threats the business faces.

## INTRODUCTION

Analytics is only as good as the decisions that it enables. The key to making the right decisions is ensuring that the right information is given to the right people at the right times.

Environments in which analytics are used can often be highly regulated. This can result in the necessity to be able to document and record the processes that were followed in handling data and the decisions made from the output of the analytics performed.

This paper will use an Insider Threat solution as the focus of how workflow can be applied and used to help direct the decisions that are made in relation to the outputs of analytics.

The paper will use the business problem of Insider Threat as an example of a solution that can use SAS Visual Investigator, which will be able to use the workflow functionality to enhance the offering. It will look first at what an insider threat is and how threats are identified. It will then discuss the workflow capabilities of SAS Visual Investigator and talk about how this can be applied to the insider threat business problem.

## WHAT IS INSIDER THREAT?

The last 10 have seen an emergence of cyber-attacks and security measures being put in place by both organizations and individuals to try to prevent insider threats. This ranges from businesses putting firewalls and implementing two factor authentication right down to individuals installing anti-virus software to protect their PCs at home.

It is predicted that the cybercrime will cost the world $6 trillion a year[1] annually by the year 2021. In addition, it is predicted that over $1 trillion will be spent globally on cyber security over the next four years.

The statistic that might surprise a lot of people in relation to cybercrime is that 60%[2] of all cybercrime is perpetrated by individuals who are known to the organizations involved. A lot of people associate cybercrime with the high profile hacking cases that have taken place in recent times.

[1] http://cybersecurityventures.com/cybersecurity-market-report/

[2] http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&htmlfid=SEW03133USEN&attachment=SEW03133USEN.PDF

Some really high profile instances of insider threats that most people will remember include the following:

- Edward Snowden, the founder of WikiLeaks, is quite possibly the best known case. He was a former CIA employee and a former US government contractor who stole and leaked classified information from the National Security Agency (NSA).
- The Panama papers are another more recent example. The employee who leaked this information has never been publicly identified[3]. He cited income inequality as his reason for leaking the papers.

Not all insider threat cases are malicious. Approximately one third of the incidents of insider threat were carried out by inadvertent actors. These are individuals who unwittingly allow access to corporate data and networks by providing access to third parties or by not following security procedures.

The other side of this is the malicious insiders. This could be employees, contractors, or anyone with access to internal data and systems. These individuals purposely set out to steal data or to cause harm to the associated organization.

*"In 2017, the insider threat epidemic begins"[4] - James Scott*

A recent Institute for Critical Infrastructure Technology (ICIT) study highlighted that the threat faced by organizations from insiders was likely to grow over the coming years. For example, terrorist organizations are expected to try to radicalize airport employees as a means to enabling terrorism.

As these threats grow, organizations will be forced to invest more in trying to identify threats as quickly and efficiently as possible.

## USING ANALYTICS TO DETECT THREATS

SAS Analytics for insider threat deterrence is based on the following four distinct analytic domains that look at data from different perspectives:

- Rules test all behaviors and activities against a predefined set of algorithms or business rules that can detect known types of risk behaviors based on specific patterns of activity or defined actions.
- Anomaly detection, such as clustering techniques, determines baseline behaviors for both individuals and groups and patterns of activity to define what's normal for each, measuring variations from the norm.
- Predictive modeling and data mining uses historical data or large amounts of transactions to predict future behavior and potential risks, as well as to detect new or emerging threat behaviors.
- Network or link analysis goes beyond data visualization to calculate the statistical significance between connections or transactions in the data and determines inferred relationships.

The combination of powerful data aggregation, a hybrid analytical approach, and a powerful technology platform enables organizations to assume a more proactive and contextually aware security posture. This approach is critical to circumventing a potential terrorist plot, thwarting an espionage mission, reducing fraud transactions, and addressing other complex threats.

All of the approaches above can be used to trigger alerts whenever a threat is detected. These alerts can be fed into SAS Visual Investigator to allow an analyst to triage them. In triaging the alert, the analyst will be able to make a decision as to whether the alert merits further investigation. If further investigation is required, then the analyst can instantiate the creation of a case.

---

[3] https://en.wikipedia.org/wiki/Panama_Papers

[4] http://icitech.org/icit-brief-in-2017-the-insider-threat-epidemic-begins/

## WORKFLOW WITHIN SAS VISUAL INVESTIGATOR

SAS Visual Investigator introduces the ability to add workflow to the cases that are modeled within a solution.

The workflow functionality is based on the industry standard Business Process Model and Notation (BPMN). Whilst not yet supporting all of the BPMN tasks, events, and gateways, it provides enough capabilities to enable the modeling of almost all of the workflows that will be required for solutions on top of SAS Visual Investigator.

### BPMN COMPONENTS

SAS Visual Investigator supports the following BPMN tasks, events, and gateways:

- Start Event
- User Task
- REST Service Task (custom to SAS Visual Investigator)
- Script Task
- Exclusive Gateway
- End Event

### Start Event

Each workflow within SAS Visual Investigator will have one start event (see Figure 1). In the process of starting the workflow it is possible to pull values from the associated case and store these values within process variables. For example, if the case has the value “High” in a risk column, then this can be passed in to the workflow and stored within a process variable. Process variables are used within conditions that control the direction the workflow takes and the tasks that result from this.

**Figure 1. A Start Event**

### User Task

A user task (see Figure 2) within a workflow is something that is assigned to either a group or a user within SAS Visual Investigator. If the task is assigned to a group, then members of that group will be able to claim it. Claiming tasks prevents multiple members of the group from working on the same task at the same time.

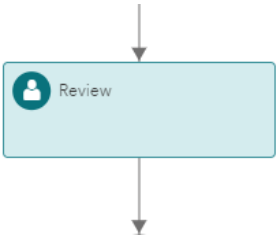



**Figure 2. A User Task**

When adding a user task to a workflow, the designer will be able to set the options that will be presented to the user performing the task to mark it as completed. For example, if the task is a review task then the user might be presented with the following options:

- Accept
- Reject

Associated within each of the options will be some configuration that will set both process variables and values within the associated case. Using the example above, if the user selects the Accept option then this might result in a column within the case being set to "Accepted". In addition, it might result in a process variable being set to Accepted. This variable might be used within a subsequent exclusive gateway to control the direction of the workflow.

## Rest Service Task

REST service tasks (see Figure 3) can be used to allow the workflow to interact with external systems. This allows process variables to be sent to the service being called through either the query string or the body of the message. The use of the REST service task is a powerful feature to allow for SAS Visual Investigator to be integrated into an organization's enterprise.

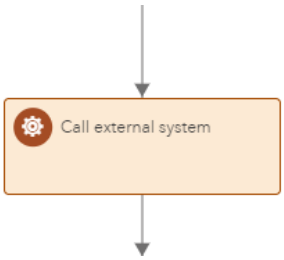



**Figure 3. A REST Service Task**

## Script Task

Script tasks (see Figure 4) allow the workflow designer to incorporate some custom JavaScript code. The normal use of this would be to manipulate and set process variables that are used to control the flow of the process.

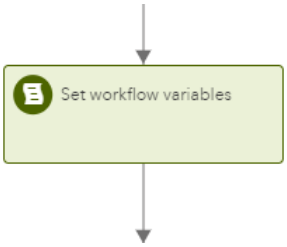



**Figure 4. A Script Task**

## Exclusive Gateway

Exclusive gateways (see Figure 5) can be thought of as “if” conditions within a workflow. They allow conditions to be created that will be used to determine the direction the workflow takes. Only one arrow out of an exclusive gateway will be followed.

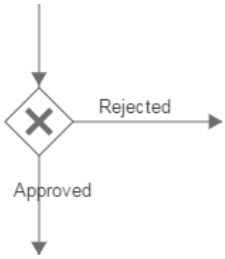



**Figure 5. An Exclusive Gateway**

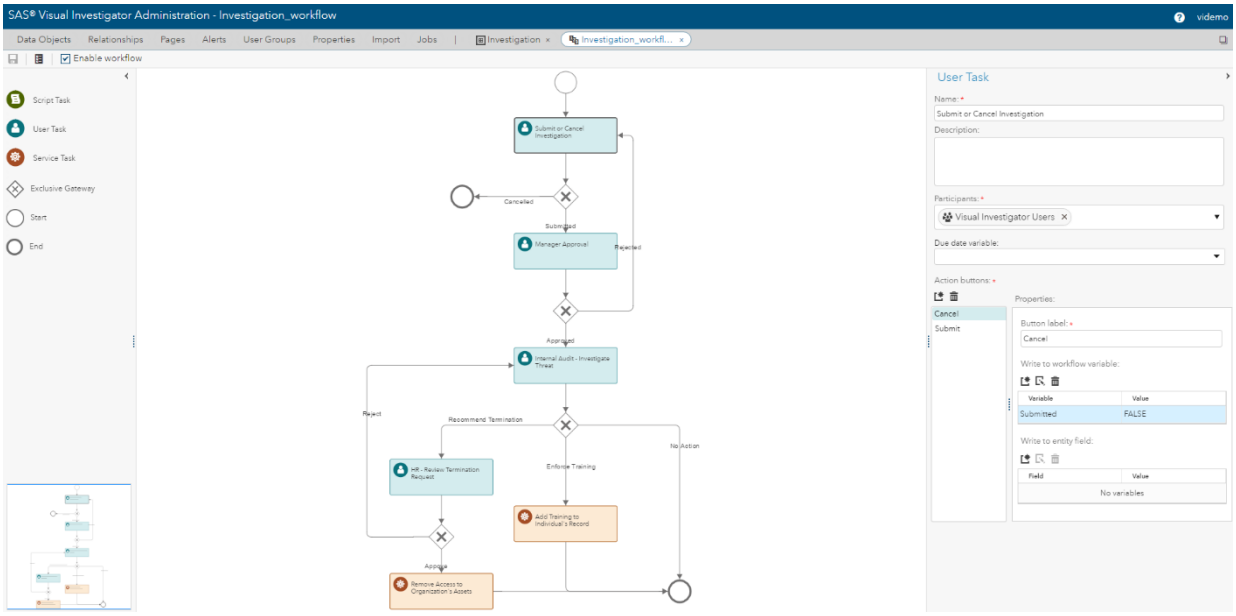
## End Event

A workflow can have multiple end events. These end events (see Figure 6) denote that the workflow has reached a point at which it can be thought of as complete.

**Figure 6. An End Event**

## ADMINISTRATION

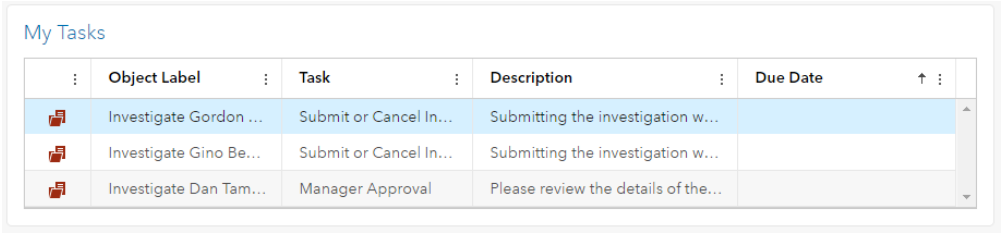
The Administration section of SAS Visual Investigator now provides access to a workflow designer (see Display 1Display 1. Workflow Designer within SAS Visual Investigator). This access has been integrated into the product to allow for tight integration between the workflow and the underlying data. Values need to be pulled from the entities into process variables within the workflow. The flip side is that the workflow needs to be able to write back to the associated entity to allow for states to be set.




**Display 1. Workflow Designer within SAS Visual Investigator**

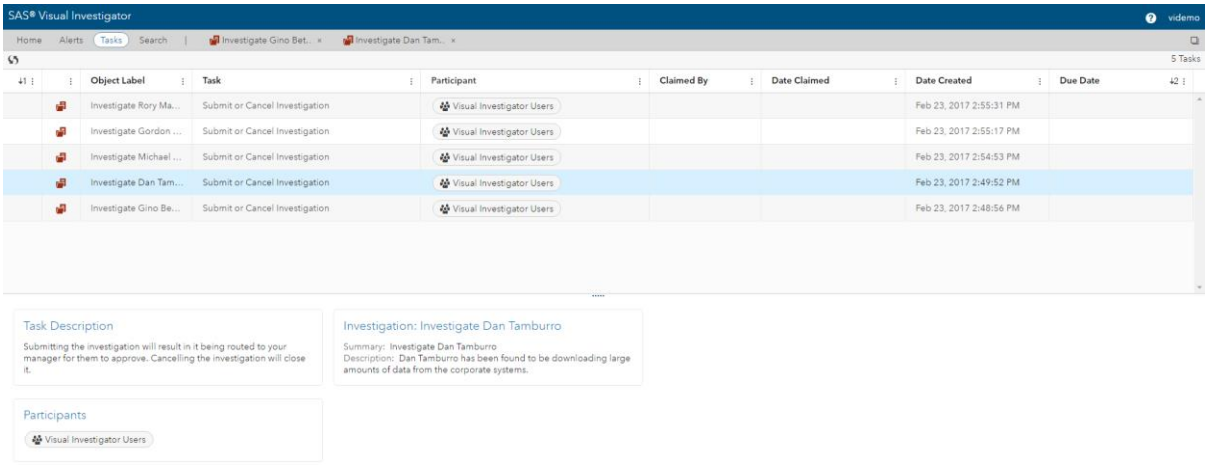
## ACCESSING WORKFLOW TASKS

There are two ways that a user of SAS Visual Investigator can access their tasks. The first way is through the homepage. A new homepage control has been added that allows users to quickly access the tasks that they have claimed (see Display 2).Display 2. My Tasks Section on Homepage




**Display 2. My Tasks Section on Homepage**

The second way that users can see the tasks is through the new Tasks tab. The tasks tab allows a user to see all of the tasks that are assigned directly to them, or to a group to which they belong (see Display 3).

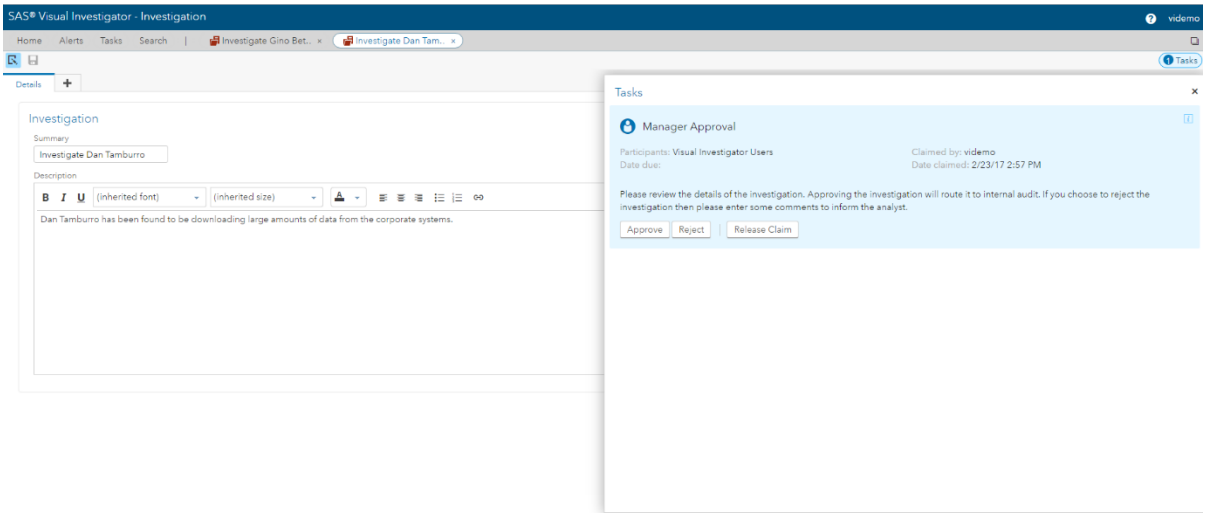



**Display 3. Task Listing within SAS Visual Investigator**

Selecting a task within the grid allows the user to see some details about it in the pane below. This includes a description of the task to be performed along with some details of the associated object.

## COMPLETING WORKFLOW TASKS

Opening an object that has an associated workflow task will now result in a new toolbar button appearing. This button shows the count of tasks that are currently outstanding on the object. If the user clicks on the button, then a pane will slide in showing details of the tasks (see Display 4).




**Display 4. Tasks Pane When Viewing an Object**

From within this task pane, the user will be able to claim the tasks. Once the tasks have been claimed, the options for completing the task will be available.

## APPLYING WORKFLOW TO AN INSIDER THREAT SOLUTION

Now that we have an understanding of the workflow capabilities of SAS Visual Investigator, we can start to look at how this could be applied to the Insider Threat solution.

As mentioned earlier, SAS uses a hybrid analytical approach to generate alerts on employees and contractors within an organization. These alerts are routed to analysts to triage and potentially investigate further.

If an analyst decides that an alert is worth further investigation, then they will create an investigation object. The creation of the investigation will result in a workflow being instantiated (see Figure 7).

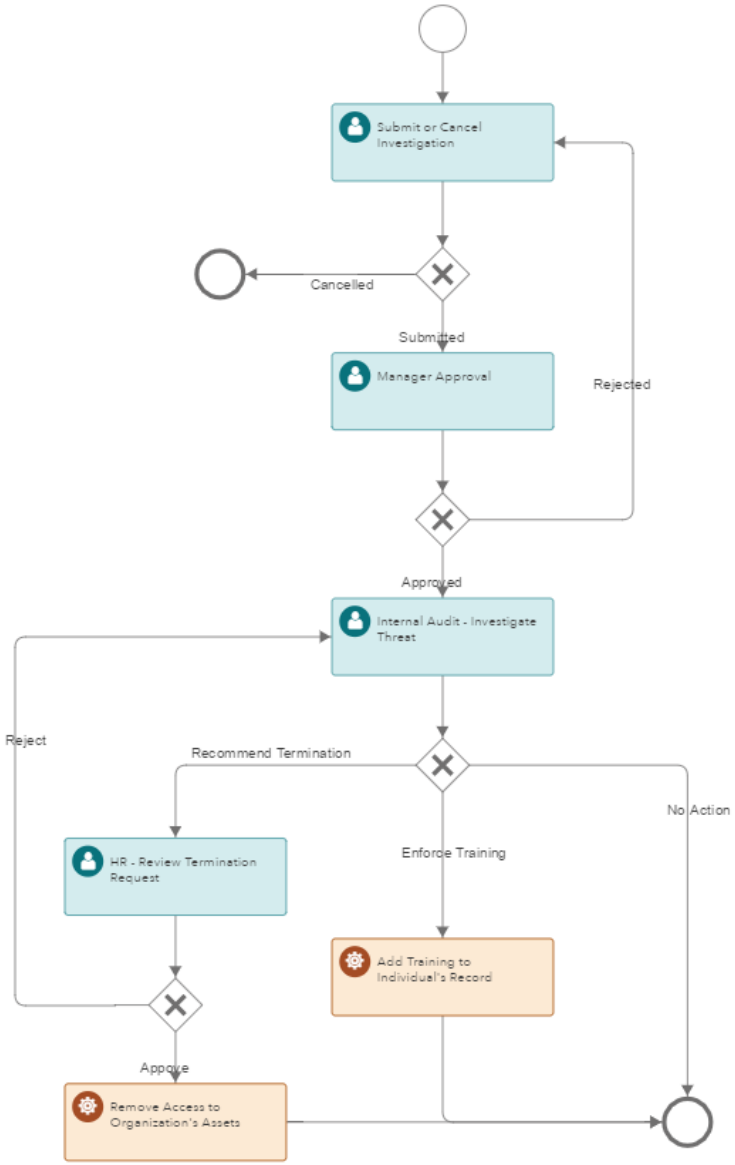



**Figure 7. Insider Threat Workflow**

The investigation object allows an analyst to document their findings into the possible threat. The analyst can use the insight capabilities of SAS Visual Investigator to do so. This allows them to snapshot any visualizations that they have generated, insert any images they might have found and to add supporting text.

Once the analyst has completed the documentation of their findings, they will be able to submit the investigation to their manager for review. The manager would have the option of approving or rejecting the investigation.

If the manager rejects the investigation, then they will fill in details of why they have done so and send it back to the analyst. The analyst then has the option of entering more information into the investigation and resubmitting it, or to cancel it. Canceling the investigation would result in the workflow being completed and therefore no further tasks being required. Submitting it again with more information would result in the manager reviewing it again.

If the manager approves the investigation, then it is routed to the Internal Audit team. The Internal Audit team is responsible for performing deeper investigations into the actions of individuals within the organization. This team will look at the information provided to them by the analyst and make a decision on how to move forward.

There are a number of options open to this team in relation to the investigation are the following:

- Recommend Termination – In this case the investigation is routed to the HR team. Once the HR team confirm the termination, a service task is used to automate the removal of access to the organization's resources.
- Require Training – If the case is found to not be malicious then the auditor might enforce that the individual involved takes a training course to make them aware of corporate security policies. This would use the service task capabilities to automatically add the details of the training to the individual's records.
- No Action – The auditors might find, after further investigation, that there was justification for the actions of the individual and that no further action is required.

## CONCLUSION

The workflow functionality within SAS Visual Investigator supplements its alert triage capabilities by allowing resulting cases to be created and managed. Having both of these functions handled within a single application reduces the cost and support burden that comes with running multiple applications.

As we have seen from the example of the Insider Threat solution, the workflow capabilities of SAS Visual Investigator allow it to manage the tasks associated with cases along with allowing it to integrate with other systems by making external REST calls.

SAS plans to build on the workflow capabilities of SAS Visual Investigator in the future by extending the support for BPMN tasks, events, and gateways.

## REFERENCES

SAS. 2017. "Using Analytics to Proactively Deter Insider Threats." Accessed February 14, 2017. https://www.sas.com/en_us/whitepapers/using-analytics-to-deter-insider-threats-107092.html.

CyberSecurity Ventures. 2017. "Cybersecurity Market Report." Accessed February 14, 2017. http://cybersecurityventures.com/cybersecurity-market-report/.

IBM. 2016. "Reviewing a year of serious data breaches, major attacks and new vulnerabilities." Accessed February 14, 2017. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&htmlfid=SEW03133USEN&attachment=SEW03133USEN.PDF.

Wikipedia. 2017. "Panama Papers." Accessed February 14, 2017. https://en.wikipedia.org/wiki/Panama_Papers.

ICIT. 2017. "ICIT Brief: In 2017, The Insider Threat Epidemic Begins." Accessed February 14 2017. http://icitech.org/icit-brief-in-2017-the-insider-threat-epidemic-begins/.

## RECOMMENDED READING

- *SAS® Visual Investigator 10.2: Administrator's Guide*
- *SAS® Visual Investigator 10.2: User's Guide*

## CONTACT INFORMATION

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

Gordon Robinson
SAS Institute Inc.
+1 984 789 7548
gordon.robinson@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.

Paper 611-2017

# Counter Radicalization through Investigative Insights and Data Exploitation Using SAS® Viya™

Lawrie Elder, SAS Institute Inc.

## ABSTRACT

This end-to-end capability demonstration illustrates how SAS Viya can aid intelligence, homeland security, and law-enforcement agencies in counterterrorism activities. Many of us are familiar with recent examples of agency failure to apportion significance to isolated pieces of information that, in context, are indicative of an escalating threat, and that require intervention. Recent terrorist acts have been carried out by radicalized individuals who should have been firmly on the organizational radar. Although SAS products perform analysis and interpretation of data that enables the law enforcement and homeland security communities to recognize and triage threats, intelligence information must be viewed in its full context. SAS Viya can rationalize previously disconnected capabilities in a single platform, empowering intelligence, security, and law enforcement agencies. SAS® Visual Investigator functions as a hub for SAS® Event Stream Processing, SAS® Visual Scenario Designer, and SAS® Visual Analytics, combining network analysis, triage, and, by leveraging the mobile capability of SAS, operational case management to drive insights, leads, and investigation. This hub provides the capability to cross-reference both internally and ingested externally held data and, crucially, operational intelligence gained from normal policing activities. This presentation chronicles the exposure and substantiation of a radical network and describes tactical and strategic disruption.

## INTRODUCTION

This paper begins with an overview of the terrorist threat currently faced by the European and Transatlantic communities and then explains how SAS capabilities can support agencies engaged in counterterrorism efforts. While acknowledging that the origins and nature of current terrorism threats vary significantly, this paper focuses primarily on the Salafist (radical Sunni) extremist threat.

### Environment and History

Recent military success against Salafi jihadist terrorist groups has seen the Islamic State in Iraq and the Levant (ISIL) losing their foothold in territories that have been considered the heartlands of Iraq and Syria. A notable consequence of these developments has been a strengthening of these groups' commitment to target Europe and North America. This approach has met with some success, drawing on the experience of European ISIL fighters returning from the frontlines. Simultaneously, they have set out to motivate "lone-wolf" activities by developing localized networks of extremists through the use of propaganda.

These ISIL activities have driven several recent high-profile, high-casualty attacks, primarily against European civilian targets in heavily populated public areas. Their tactics have varied greatly, ranging from sophisticated, highly coordinated attacks to crude, blunt-force strikes. Perhaps more significantly, these attacks have served to highlight failings in the local, national, and international intelligence and enforcement services who are perceived to have missed opportunities to preemptively disrupt them.

The changing nature of the terrorist threat has required intelligence and enforcement agencies to shift their focus and adjust their tactics. Perhaps the greatest influence on this change has been that recent attacks have been largely perpetrated by individuals who have subsequently been revealed to be known criminals. Indeed, many of ISIL's successes can be directly attributed to this ability to radicalize individuals whose history has previously been marked by petty crime.

These factors are driving changes in the counterterrorism dynamic and have exposed weaknesses in the traditional capabilities around gathering, exploiting, and sharing of intelligence within and between agencies and nations. While counterterrorism has traditionally been the domain of intelligence and homeland security agencies, recent terrorist attacks (born out of and planned within criminal networks) have to a large extent ranged beyond these services' purview. This change in dynamic has placed law enforcement at the center of counterterrorism endeavors and has, as a direct consequence, seen general policing or community information elevated to being among the most critical of data sources.

## Examples

***Sophisticated and Coordinated***: In November 2015 a brutal, highly coordinated attack took place in Paris when ISIL-inspired terrorist cells, using assault rifles and wearing suicide vests, simultaneously attacked multiple soft targets, including the Bataclan Theatre, where 89 died. The terrorist network behind the attack was led by Salah Abdesalam, a radicalized individual with strong links to known criminal networks. The group also included individuals who had previously fought in Syria.

Michael Leiter, former director of the United States' National Counterterrorism Center, commented afterward that "the attacks demonstrated a sophistication not seen in a city attack since the 2008 Mumbai attacks, and would change how the West regards the threat" of terrorism generally.

***Blunt Force:*** In July 2016 Mohamed Lahouaiej Bouhlel drove a lorry into a Bastille Day celebration in Nice, France, killing 84 people. This blunt-force attack might have lacked the sophistication and planning of the Paris attack, but it ultimately had a similarly deadly effect. Although Bouhlel was known to law enforcement for involvement in petty criminality, there were no reports of his having any direct links with a terrorist group. However, he was subsequently described by ISIL as a "soldier of Islam." Significantly, he was known to have psychiatric problems, a characteristic increasingly common in these incidents.

The evidence indicated that Bouhlel had been radicalized by ISIL propaganda, and he was subsequently classified by elements of the mainstream media as a "lone-wolf" actor. Nevertheless, he did not act alone in the planning and development of his attack, and his actions were facilitated by criminal contacts through which, among other activities, he procured a firearm.

## Future

Transatlantic law-enforcement communities have publicly acknowledged their current weaknesses and their vulnerability to future terrorist attacks. This recognition has resulted in a number of initiatives to assist with building understanding of possible ways to mitigate such threats in the future.

A significant body of work is to be found in the GLOBSEC Intelligence Reform Initiative (GRI), which recently published a paper, "Reforming Transatlantic Counter-Terrorism". One of the important observations of this paper was the following:

> *"The key problem the Globsec Intelligence Reform Initiative addresses is that of intelligence and personal data sharing and its operationalisation at the domestic as well as transnational level. Although many intelligence agencies have been at the centre of counter-terrorism efforts since 9/11, this report recognises that as terrorism is fundamentally viewed as a crime in both Europe and North America, law enforcement is increasingly at the centre of better pan-European and transatlantic counter-terrorism cooperation. Crucially, better fusion of intelligence processes, and intelligence and law enforcement agencies, is needed to provide the means for pre-empting terrorist attacks before they occur, rather than relying on effective investigation after the event."*

While the need for data sharing and the operationalization of intelligence products is widely accepted, there is also a recognition that to be effective, agencies must enhance the information technology capabilities around collation, analysis, and the associated management of operational processes.

The related significant challenges are often magnified rather than lessened by the volume of data that exists for agencies to exploit. Information sources are vast and varied, a complexity that is only increased by this now essential inclusion of day-to-day community and policing data.

## SAS VIYA PLATFORM

Supported by SAS Data Management services, the SAS Viya platform can help law enforcement, security, and intelligence agencies to address the many challenges associated with counterterrorism. SAS Viya comprises solutions built with embedded analytical capabilities at their core, allowing the exploitation of information through alerting, triage, enrichment, and operationalization. The open architecture of SAS Viya also ensures that operatives at all levels within participating organizations (including executives, analysts, investigators, and front-line officers) are always able to access their data and related insights in the most effective and relevant manner and are not tied to a single application or device.

With SAS Viya, it is possible to integrate all aspects of the intelligence and investigation life cycles through standard, unified components that provide a foundation for sharing and communicating. The major components of such a system to handle data for intelligence purposes would include (but not be limited to) the following SAS solutions:

- SAS® Visual Analytics
- SAS® Event Stream Processing
- SAS® Mobile Investigator
- SAS® Visual Investigator

### SAS VISUAL ANALYTICS

#### Business Challenge

In tackling the terrorist threat, law enforcement, security, and intelligence agencies must use their finite resources to the best possible effect. Decisions must be based on accurate assessments, and the strategic direction must always be made clear and be justifiable.

The development of an effective Strategic Assessment is dependent on skilled operatives' undertaking detailed research and analysis of all available information sources. To develop this "big-picture" document and truly understand the nature and level of the threats, agencies should not restrict their information sources to only those that are routinely maintained or accessed in the course of day-to-day operations. External influences, such as information about public perceptions, health, welfare, and education, must also be taken into consideration, as such factors can provide valuable insight into the fears, vulnerabilities, and threats extant in local communities.

By making a comprehensive and complete assessment available, agencies are better able to set a strategic direction, prioritize, make defensible decisions, and allocate resources intelligently, fully considering the operational options available to them. Counter-radicalization is a related priority, and agencies must take all necessary actions to understand the terrorist ideology, identify those who promote it, and prevent people from being drawn into terrorism. Tactical options when seeking to prevent radicalization or to preempt a terrorist attack would include proactive investigation, surveillance, education, and engaging with sectors and institutions where the risks of radicalization are greatest.

This practice of centering activity on a strategic assessment is universal among the European and Transatlantic law-enforcement, security, and intelligence agencies, and this commonality facilitates collaboration on an interagency, national, and international basis. There are many examples where this type of joint assessment has been adopted: for example, a joint endeavor of the European Council to develop the EU Counter-Terrorism Strategy[1]

[1] http://www.consilium.europa.eu/en/policies/fight-against-terrorism/.

## Applicable SAS Viya Module

SAS Visual Analytics supports agencies in creating, sharing, and acting upon interactive and meaningful intelligence products, such as strategic assessments.

By using SAS Visual Analytics, analysts gain the ability to explore the corpus of information available within the organization as well as that shared through collaboration. Products will include the vital and overarching strategic assessment, together with tactical reports reflecting priorities and supporting ongoing operations.

Given the dynamic nature of the terrorist threat, the interactive features of SAS Visual Analytics reports are crucial. These reports enable recipients to focus on the information facets that are most appropriate and in whatever manner is most relevant to the task at hand, using filters and drill-through capabilities to further explore the data and develop insights.

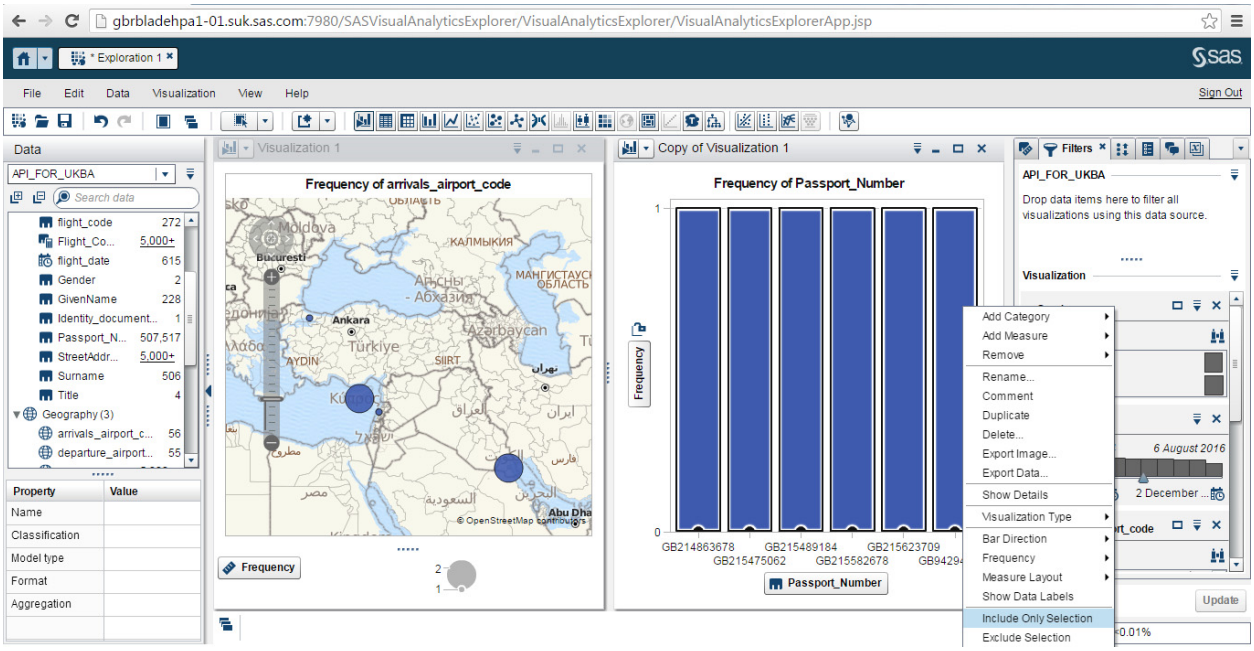



**Figure 1. Example of SAS Visual Analytics dashboard**

A standard response to a terrorist incident will see different levels of commanders taking control of the various aspects of activity (for example, overall command, referred to as Gold; tactical: Silver; and operational: Bronze). SAS Visual Analytics gives commanders easy access to explore dashboards and reports to aid in their decision-making process. The ability to access this information from mobile devices is of particular importance to bronze commanders who are often required to operate from the field (for example, taking responsibility for hostage situations or bomb scene management).

## SAS EVENT STREAM PROCESSING

### Business Challenge

There is an expectation--however unrealistic--that law enforcement and intelligence agencies have the capability to manage and exploit (at least in some form) **all** of the information sources held by or made available to them. However, even within a single organization data management can be challenging, as disparate information sets are often held in discrete “silos” – with different schemas, access rights, and organizational practices. Multi-agency collaboration only serves to increase this potential complexity, leaving practitioners with the task of interpreting significant quantities of ever-changing information, presented in a variety of ways. By adding layers of third-party, high-volume data sources (such as communications data or automatic license plate recognition data), the challenge only increases almost exponentially.

The limitation of software tools that have previously been available to agencies is that they may only be able to exploit available information after the event. An analysis of circumstances surrounding recent terror attacks would seem to indicate that investigators were unaware of critical information that was already held by their organization, which might therefore be seen to have missed opportunities for preemptive action. Inevitably, starting an investigation after an event has occurred leaves analysts and investigators playing catch-up as they try to keep pace with new investigative streams and evolving events.

### Applicable SAS Viya Module

SAS Event Stream Processing can support agencies in addressing the challenges presented by the attempt to keep up with such potentially vast quantities of information by applying analytics to the data as it becomes available.

With SAS Event Stream Processing, huge volumes of data streaming in real time from multiple jurisdictions, organizations, and nations can be filtered, categorized, aggregated, and cleansed before being stored, saving operatives from having to sort through and interpret disconnected and often polluted data sources.

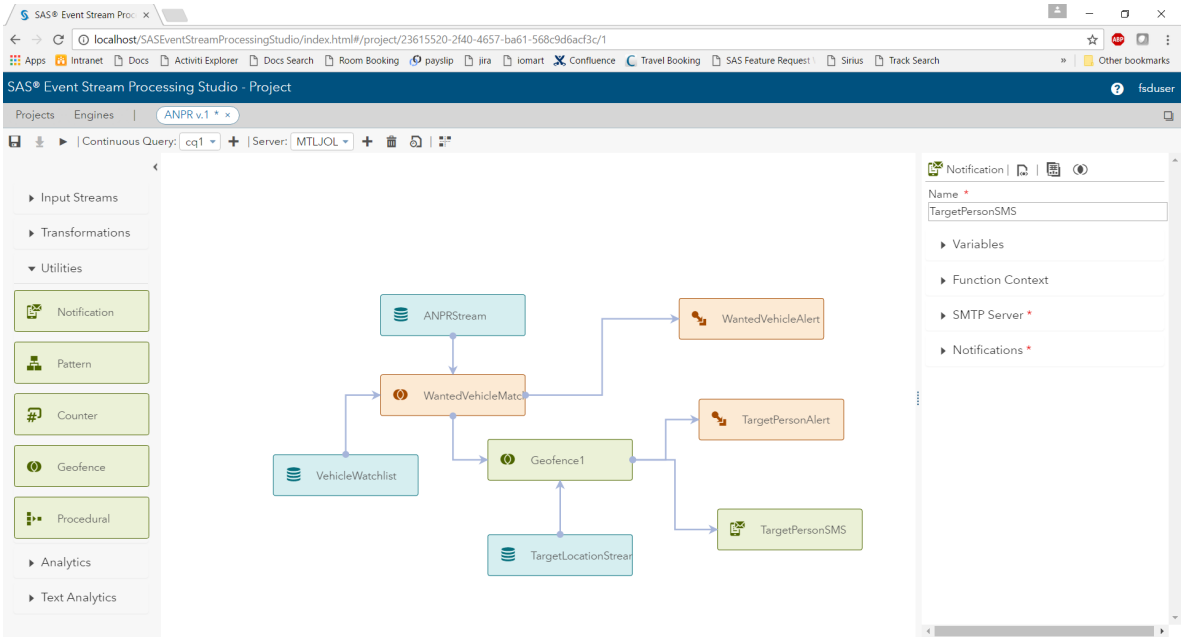



**Figure 2. SAS Event Stream Processing identifying vehicle on watch list**

SAS Event Stream Processing is a powerful tool that is capable of enhancing an organization's capacity to respond to emerging threats and take preemptive action. It can apply analytical models simultaneously to both fast-moving and static data, ensuring that relevant information is isolated and that analysts receive timely alerts related to significant events; identified criminal networks and activity; or anomalous behavior.

As an example: An alert is generated by two seemingly unconnected individuals traveling separately to a country with known affiliation to terrorism, their travel being paid for using the same credit card.

While alerts to items of significance are of great value, organizations are further challenged with converting such insights into operational action.

## SAS MOBILE INVESTIGATOR

### Business Challenge

The previously discussed acceptance that community and general policing data is now essential to the counterterrorism effort has exposed the weaknesses in existing systems. In most nations, the responsibility for the individual facets of such "day-to-day" policing is managed by distinct departments or units, potentially

even split across multiple agencies or organizations (for example, road traffic, community policing, or criminal investigation units). An unfortunate--but natural--consequence of this reality has been that vital data is held in disparate stores, and with the limitations of legacy software, there is often no simple means to search across, rationalize, or identify information of significance within these “siloed” repositories.

While this situation is generally most prevalent among law enforcement agencies, similar architectural and functional challenges exist within the wider intelligence and security communities. The need for the modernization of systems is widespread, as these agencies seek to increase their access to and their exploitation of the data available to them.

Of particular note is a recognized need to improve the ability to obtain intelligence as quickly as possible (“fast time intelligence”) in the aftermath of a significant event. Weakness in this area was clearly evidenced in reviews of the post-incident responses of law-enforcement agencies to many of the recent European attacks. And while there is no argument that officers responding to those incidents deserve the highest praise, their actions would undoubtedly have been hampered by inevitable delays in identifying crucial investigative leads within such disconnected information stores.

## Applicable SAS Viya Module

SAS Mobile Investigator is designed to meet the specific needs of intelligence, enforcement, and investigative agencies. It provides a comprehensive operational environment capable of supporting the nuanced processes of intelligence and law-enforcement agencies–an environment that is essential to ensure legislative and regulatory compliance–through key capabilities such as advanced search; tasking; operational reporting; a robust configurable security model; and comprehensive auditing.

SAS Mobile Investigator is a web-based application that uses responsive design to alter the ways that the various components are presented, ensuring that all functionality is easily available on whatever type of device is used to access the system (for example, mobile, desktop, and so on). Further, this design enables the capabilities of each device to be used as appropriate. For example, a user accessing the system via a mobile phone would be able to use the GPS capabilities of the device to log the precise location of an incident, and use the camera to capture an image or video, which could be immediately uploaded and made available.

While seemingly straightforward, the value associated with this type of mobile access cannot be overstated. Officers in the field can receive tasks, comply with due process, and upload what could prove to be invaluable information without having to return to an office or to a vehicle.

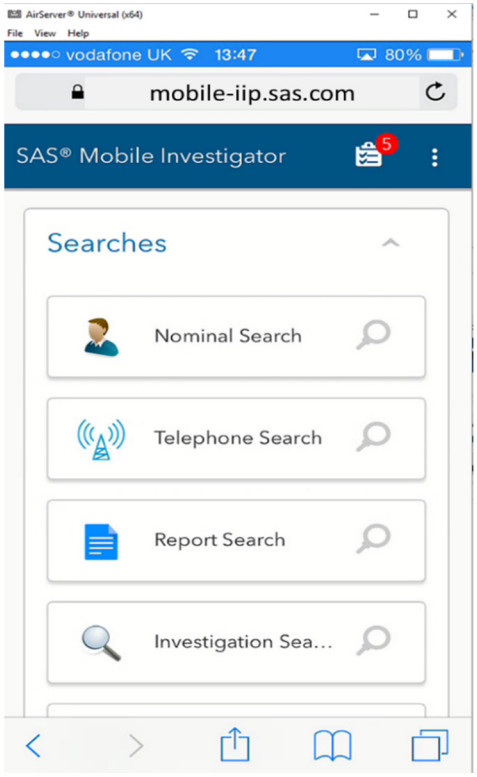



**Figure 3: SAS Mobile Investigator**

The enabling of officers to feed “street-level” intelligence directly into the corpus of knowledge about a particular individual, group, or community will enhance the agencies’ ability to spot behavioral patterns and anomalies, perhaps indicative of changing social dynamics, and to prioritize appropriate intervention, for example, investigation, education, or disruption.

Having field access to the totality of organizational data allows officers (in near real time) to review information relevant to live incidents, assess risks, and customize their responses appropriately while remaining cognizant of the "bigger picture". While these capabilities are clearly important for ongoing investigations and routine operational activity, the ability to facilitate fast time intelligence gathering and exchange of information in the immediate aftermath of a terrorist incident could prove crucial in facilitating early arrests, or preempting further attacks.

## SAS VISUAL INVESTIGATOR

### Business Challenge

The working practices of intelligence and law enforcement communities have evolved over many years and are generally well defined, reflecting the needs and priorities of individual organizations and their practitioners. These practices might still be valid today, but they must now be applied against the modern environment where the volume, variety, and velocity of data have reached unprecedented (and continually increasing) levels. In meeting these challenges, agencies require modern analytical tools that are able to refine and offer focus on relevant data while also supporting existing operational practices, so essential for intelligence development, for information sharing, and for ensuring the integrity of evidence collection.

Successful outcomes are often dependent on the early identification of factors that signify risk and require prioritization. These could include, for example, patterns within the data identifying known criminal networks, or a collection of (sometimes related; sometimes seemingly disparate) elements that indicate escalating risk. A real life example relating to the Salafi jihadist threat would be a pattern of actions, travel, communications, and lifestyle changes known to be a precursor to radicalization.

### Applicable SAS Viya Module

SAS Visual Investigator provides an environment where operatives can–through the processes of alert generation, search, and the application of advanced analytics--work on and keep pace with the volume and variety of data that is now available to be exploited. Similar to SAS Mobile Investigator, SAS Visual Investigator supports (and where required, enforces) the nuances of operational process that are required for legislative and regulatory compliance.

As the nature of terrorist threats evolves, agencies will be required to adjust their focus to seek out objects and patterns within their data that could be of significance and require action. Alert generation through the application of rules and algorithms can highlight items of interest and, where possible, support early intervention.

Importantly, the graphical scenario builder feature in SAS Visual Investigator gives agencies the flexibility to address changing threats by designing, testing, and iterating rules that can automatically generate alerts on matching patterns within the data. This process can generate an alert based on the identification of a range of factors that, in isolation, might seem unrelated or of little significance, but when viewed as a whole, could be indicative of an escalating risk. For example, analysis might identify a pattern of behavioral factors, travel, and communications that could be associated with radicalization.

Scenarios that are designed within SAS Visual Investigator can also be enhanced using SAS Event Stream Processing to generate alerts from volume data in motion.

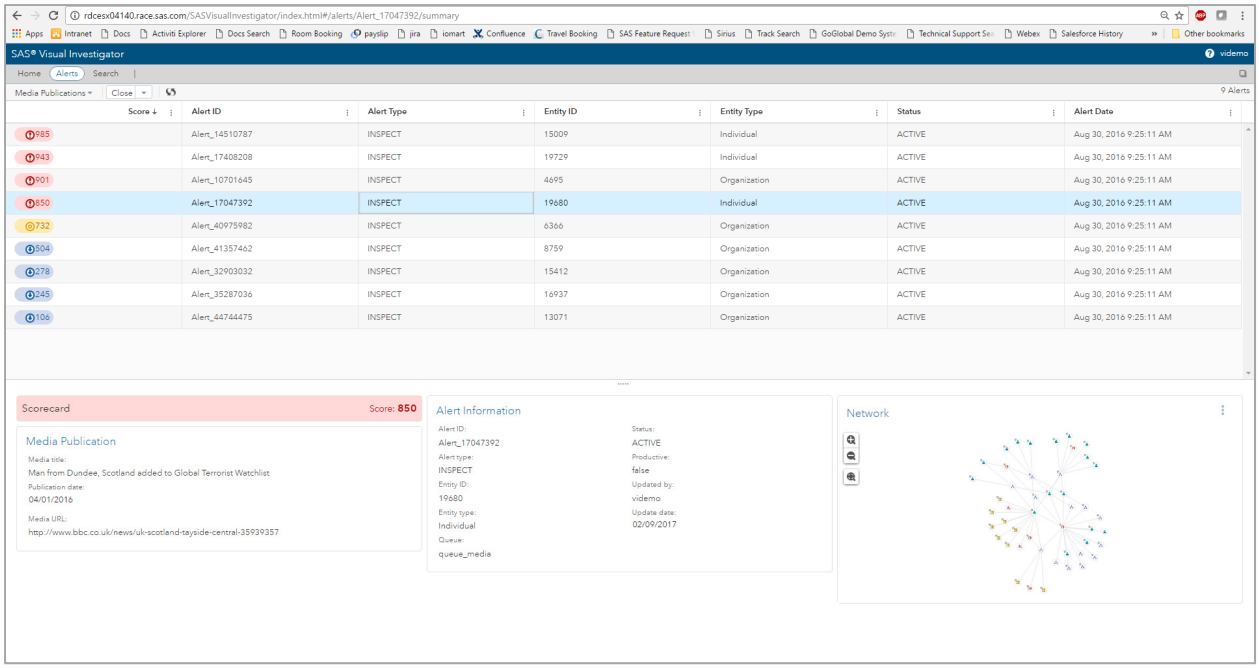



**Figure 4. Example of SAS Visual Investigator alert management dashboard**

The operationalization of data (including analytically derived alerts) is of paramount importance in counterterrorism efforts. SAS Visual Investigator supports triage, prioritization, and assignment of responsibility. Advanced analytical capabilities allow agencies to develop intelligence insights and uncover investigative streams.

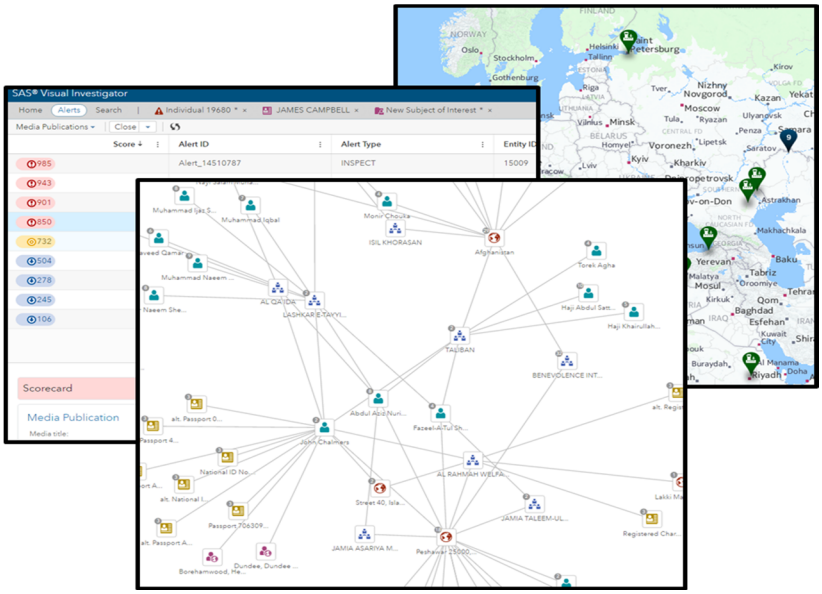



**Figure 5. SAS Visual Investigator network diagrams, map views**

In addition, insights that are derived from the data (such as network diagrams, timelines, or map views) can be used to create operational products that are essential to advance the work of agencies. For example, the data in SAS Visual Investigator can be used in the development of subject profiles, target packages, or threat assessments.

Crucially, data that is managed and developed within SAS Visual Investigator will be accessible through SAS Mobile Investigator, enabling officers to conduct research while in the field and receive tasks stemming from deskbound research.

## CONCLUSION

The European and Transatlantic intelligence, security, and law-enforcement agencies are well aware of the changes and improvements required to be successful in meeting the real and growing threat of terrorism. While significant progress has been made in international and interagency collaboration, clear weaknesses remain.

It is widely accepted that organizational disconnects can exist in all areas and at every level of an agency and are regularly manifested in ineffective communication between stakeholders and an inability to fully exploit the available information assets.

The magnitude of the counterterrorism challenge cannot be overstated. While there is no “magic bullet” to solve the problems that global terror poses, SAS Viya represents a unique opportunity to work toward the much-needed cohesion in approach and to build on the existing corporate knowledge of the threat. In a single platform, SAS Viya offers a comprehensive set of capabilities to manage huge volumes of data while simultaneously facilitating strategic and operational activities through a combination of advanced analytics and business process support.

## REFERENCES

GLOBSEC Intelligence Reform Initiative - Reforming Transatlantic Counter-Terrorism

(http://www.cepolicy.org/sites/cepolicy.org/files/attachments/giri_report_1.pdf)

European Council to develop the EU Counter-Terrorism Strategy
(http://www.consilium.europa.eu/en/policies/fight-against-terrorism/)

## CONTACT INFORMATION

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

Lawrie Elder
SAS Investigation and Intelligence Practice
lawrie.elder@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.

Paper SAS603-2017

# You Imported What? Supporting International Trade with Advanced Analytics

Susan Trueman, SAS Institute Inc.

## ABSTRACT

Global trade and more people and freight moving across international borders present border and security agencies with a difficult challenge. While supporting freedom of movement, agencies must minimize risks, preserve national security, guarantee that correct duties are collected, deploy human resources to the right place, and ensure that additional checks do not result in increased delays for passengers or cargo. To meet these objectives, border agencies must make the most efficient use of their data, which is often found across disparate intelligence sources. Bringing this data together with powerful analytics can help them identify suspicious events, highlight areas of risk, process watch lists, and notify relevant agents so that they can investigate, take immediate action to intercept illegal or high-risk activities, and report findings. With SAS® Visual Investigator, organizations can use advanced analytical models and surveillance scenarios to identify and score events, and to deliver them to agents and intelligence analysts for investigation and action. SAS Visual Investigator provides analysts with a holistic view of people, cargo, relationships, social networks, patterns, and anomalies, which they can explore through interactive visualizations before capturing their decision and initiating an action.

## INTRODUCTION

Unprecedented volumes of cargo and people are now crossing international borders. This freedom of movement is critical to supporting global trade and growing economies but places border agencies under huge pressure.

In their 2016 Annual Review, the International Air Transport Association (IATA) estimated the value of international trade shipped by air in 2015 was $5,7 trillion. And tourists traveling by air spent over $620 billion (IATA 2016, p. 14). The UN Conference on Trade and Development also estimated that, in 2015, the volume of seaborne trade exceeded 10 billion tons for the first time (UNCTAD 2016).

Each year, containers arrive in the USA through the following means (U.S. Customs and Border Protection 2017):

- 11 million maritime containers arrive at seaports
- 11 million containers arrive at land borders by truck
- 2.7 million containers arrive at land borders by rail

These numbers are staggering and set to increase. Border agencies must ensure that they maintain security, intercept passengers who might be traveling illegally, and also identify illegal or high-risk freight and contraband, even when faced with constrained budgets and increased volumes of travelers and freight arriving at ports every day.

This must be done while honoring stringent data privacy laws and minimizing disruption to legitimate cargo and passengers to ensure a country's ports remain competitive in the global market. According to the IATA, in order to improve competitiveness, the air cargo industry's aim is a 48-hour reduction in average shipping times by 2020 (IATA 2015, p. 40). This means any proactive action taken by agencies must be efficient and not introduce overheads and delays.

This paper looks at the challenges faced by border agencies and shows how the new features in SAS Visual Investigator can help by bringing together disparate data, detecting suspicious activity, presenting analysts with alerts to be triaged, and enabling investigators and intelligence analysts to conduct investigations and make appropriate, data-driven decisions.

## BORDER MANAGEMENT: THE CHALLENGES

Each country is responsible for its own border control, and national border agencies face a number of challenges, including the following:

- national security and counter-terrorism
- monitoring and managing immigration
- combating human trafficking and migrant smuggling
- preventing the import and export of contraband including drugs and counterfeit goods.
- ensuring that ports of entry are connected and competitive
- efficient use of resources and public funds
- accurate collection of duties
- compliance with local and international legislation (including privacy laws)

Physical security combined with new processes and legislation are important and can help address these challenges. However, in a world of data, it is incumbent upon border agencies to also use processes and technologies that will ensure that they are pro-active and efficient in identifying and investigating suspicious events and activities.

It is not enough to simply collect data. Organizations must make good use of their data assets to ensure that high-risk passengers and cargo shipments are intercepted without a negative impact on the processing of legitimate passengers and freight at borders.

### MAKING USE OF DATA

Many governments have increased their data gathering requirements. For example, airlines operating in some countries must collect and share Advance Passenger Information (API). The same is true for cargo as governments now look for airlines and other transportation industries to provide Advance Cargo Information (ACI).

In addition, border agencies have access to a variety of other data sources, including manifest details, requests for visas and other permits, watch list and sanctions information, sensor data, law enforcement records, intelligence reports, and so on. These huge volumes of data pose a challenge.

Agencies have finite human resources in the form of agents, analysts, and investigators. It is impossible for analysts to manually identify all unusual patterns of activity, bad actors, and other anomalies across this vast quantity of data. However, border agencies can use proactive analysis to identify high-risk cargo or passengers and make all data accessible to analysts and investigators so that they can take prompt action.

The United States Customs and Border Protection summed this up when they acknowledged their need to enhance their ability to collect, analyze, and appropriately share intelligence and information. This includes providing timely warnings of potential threats and proactive enforcement opportunities (U.S. Customs and Border Protection 2016, p. 12).

In addition to the sheer amount of data involved, the data is often stored in disparate silos. Historically analysts and investigators had to run separate queries in multiple isolated systems to find information that might be relevant to a particular individual, shipment, organization, or other item of interest. They would then have to manually examine each piece of seemingly unrelated data in an attempt to identify patterns and anomalies. Visualizing and sharing their findings relies on their ability to manually create a visual representation of the associations between entities or perhaps pin them to a map. Relying on analysts to do this unaided is error-prone and inefficient.

## HOW CAN SAS VISUAL INVESTIGATOR HELP?

To make the most of their human assets and ensure that they make prompt decisions backed up by data, border agencies must take a pro-active approach to presenting critical information to analysts and agents. They cannot rely on individuals using time-consuming, manual processes to identify high-risk passengers and cargo shipments or other suspicious activities.

SAS Visual Investigator does the following:

1. delivers a holistic view of all available data.
2. identifies previously concealed relationships and builds social networks around people, organizations, cargo shipments, and other objects.
3. enables the design of surveillance scenarios and rules to identify areas of concern around high-risk or suspicious freight, travelers, or events and automatically alerts analysts and agents.
4. enables efficient, alert triage and disposition.
5. provides analysts with the ability to search multiple internal and external data sources from single application interface.
6. helps analysts and investigators understand and explore their data in different visualizations (for example, on a geospatial map or plotted in chronological order on a timeline).
7. enables agents and analysts to conduct and manage cases and targeted investigations supported by business process and workflow.
8. enables analysts to capture their findings and the output of investigations to support their decisions and actions.
9. supports collaboration within an agency's investigative teams.

We'll look at some of these key capabilities in more detail from the point of view of three different user roles commonly found within a border agency's investigation unit: System Administrator, Analyst, and Investigator.

### ACCESSING DISPARATE DATA

Border agencies have access to a lot of data that is found not only in large volumes but often housed in different data stores that need to be accessed by many different systems. For example, separate databases might be used to store advanced passenger/cargo information and intelligence reports. The ability to bring this data together and make it accessible from a single application is built into the very foundations of SAS Visual Investigator.

SAS Visual Investigator focuses on two main categories of data:

- external entities – data that is stored in an external data store but is used by SAS Visual Investigator for surveillance, driving entity resolution and social network building, indexed for search, and made available in all end-user visualization without having to create multiple copies of the data.
- internal entities – objects created, edited, managed, and indexed within SAS Visual Investigator's primary data store. This can include investigations and cases created and edited by agents, analysts, and investigators within the SAS Visual Investigator application. This data can also be used as an input for surveillance, network building, search, and visualizations.

SAS Visual Investigator can also make use of large volumes of transactional data that is not indexed for search but is displayed in specific visualizations. Administrators can use a drag-and-drop interface to design the pages used to display data to the end-users and decide, with the click of a button, which search filters and facets to expose.

As new data sources become available, administrators can simply use the built-in administration

application to configure SAS Visual Investigator to connect to the new data stores.

## SURVEILLANCE SCENARIOS

SAS Visual Investigator provides the ability to author surveillance rules and scenarios to identify anomalies or suspicious activity and generate alerts to be triaged and investigated.

Using the visual, point-and-click interface shown below, users can design multiple scenarios and test each one on a sample of their data before publishing to the live system.

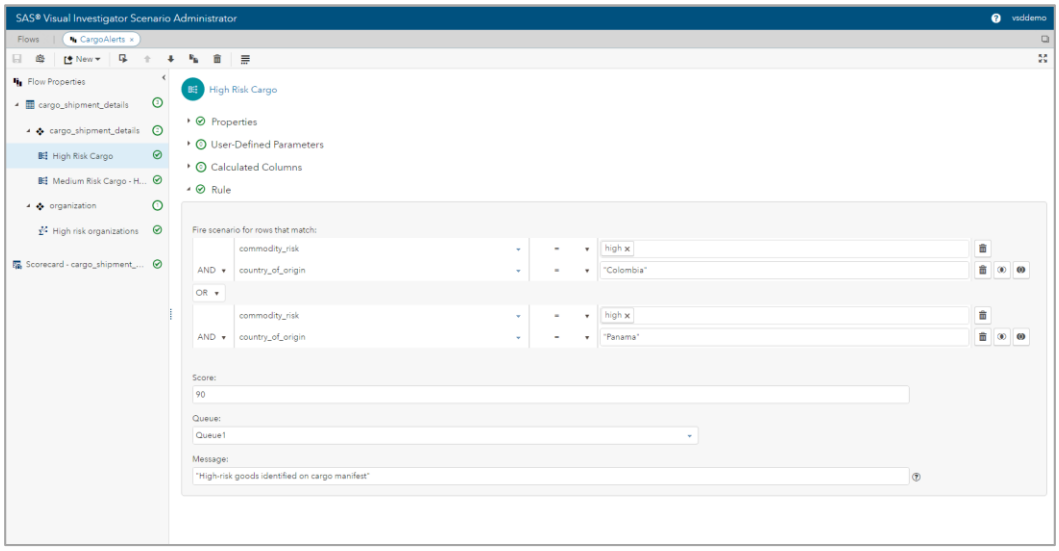



**Figure 1. Authoring Surveillance Scenarios within SAS Visual Investigator**

Surveillance scenarios assess risk and identify patterns of activity that might otherwise go unnoticed by human analysts. Scenarios can range from simple rules to complex analytical models. The output of the surveillance process is the generation of alerts.

Each alert represents a positive match against defined rules and includes details of exactly which rules were triggered as well as information about the alerted entity or subject of the alert, such as a specific passenger, freight shipment, organization, and so on.

A score is calculated to provide analysts with a clear picture of the risk associated with this event. The scores for each scenario are weighted and rolled up to an overall scorecard for a given alert. This provides analysts with a clear overview of the scenarios and rules that were triggered and ensures that the alert is routed to the correct queues and strategy.

Analysts and agents are assigned to specific strategies, and from here they can view, review, triage, and investigate each alert and take action.

As noted by U.S. Custom and Border Protection, systems must be able to adjust to changes in trade patterns and trends (U.S. Customs and Border Protection 2016, p. 24). SAS Visual Investigator gives agencies ownership and control of their surveillance scenarios. They can respond to changes in their environment by modifying existing scenarios or adding rules based on new information, analyst knowledge, and successful outcomes from previous investigations.

## ALERT MANAGEMENT AND TRIAGE

SAS Visual Investigator's alert management process is designed to help analysts and investigators be proactive and take action on the output of an analytical process or surveillance scenario.

Alerts generated as a result of surveillance scenarios are served up to analysts and investigators via queues and strategies. Each strategy within SAS Visual Investigator represents different business problems being tackled by the border agencies. Strategies help organizations allocate specialized resources effectively. For example, separate strategies can be set up for the following:

- suspected counterfeit cargo
- consignment shipped from a high-risk shipping company or port
- cargo identified as high-risk goods from a specific freight company
- high-risk passenger who booked flight within 24 hours of travel and paid cash
- cargo from suppliers with connections to organizations or individuals on sanctions watch lists

As shown below, when analysts log on to SAS Visual Investigator, they are presented with an overview of the work allocated to them, and they can review the alerts in each strategy they have been assigned. They might be responsible for monitoring and working multiple strategies or focus on one specific area of risk.

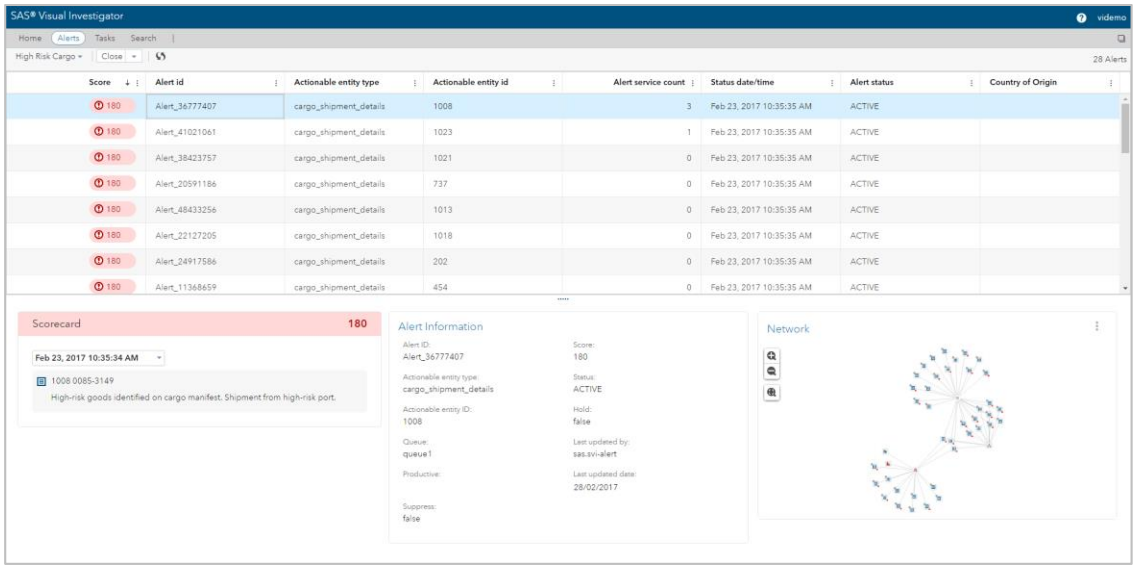



**Figure 2. Alert Triage within SAS Visual Investigator**

Within each strategy, an organization can define a number of different queues to represent the priority of the alerts. The score calculated by the surveillance process is clearly visible to the analysts as an indication of the severity and associated risk. This is combined with a detailed scorecard, an overview of all triggered surveillance rules, the history of the alert, and details of the associated entity to help analysts establish a clear picture of an event, threat, or emerging situation.

From the alert triage screens, analysts can review key information about each alert before using SAS Visual Investigator's rich visualizations to further understand the context of the alert and deciding which disposition action to apply. For example, they might delve into the associated social network or view a timeline of events before deciding to close the alert, suppress the alert or escalate to a case for further investigation. The available disposition actions will depend on an organization's specific process as they can be configured and extended by administrators.

In addition to alerts being generated with SAS Visual Investigator through the use of pre-defined surveillance rules, analysts might also raise manual alerts on other entities, such as people, vessels, aircraft, shipping companies, and so on that they identify as high-risk or suspicious during the course of their investigation.

## SEARCH, DISCOVERY, AND EXPLORATION

While system-driven surveillance and alerting is critical to ensuring suspicious events and anomalies are identified and quickly sent to analysts for action, it is also essential that analysts and investigators have the ability to follow their own hypotheses and lines of inquiry (for example, responding to an ad hoc request or tip from a partner agency).

The search and discovery functionality within SAS Visual Investigator supports the alert triage process as well as the user-driven research and investigations.

Powerful free-text and geospatial search capabilities enable analysts to search across all internal and external data sources. Search results are returned in order of relevance, and filters and facets can be used to refine the list of results as shown below.

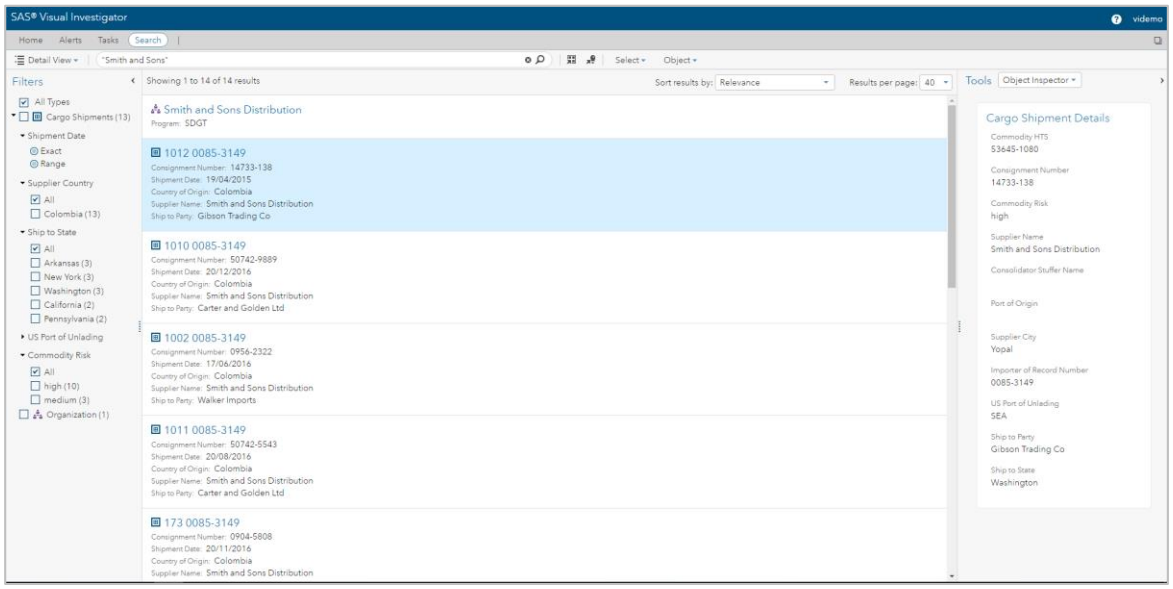



**Figure 3. Visualizing Search Results within SAS Visual Investigator**

Analysts can display their search results on a map view to display the geospatial data associated with their results set as shown below. From the map view, they can also initiate a search that combines free-text search criteria with one or more geospatial search areas.

At any time, the analysts can select specific entities of interest to be added to workspaces within alerts, investigations, or cases for further analysis at a later date, or to explore the data using additional investigative tools such as the network or timeline views.

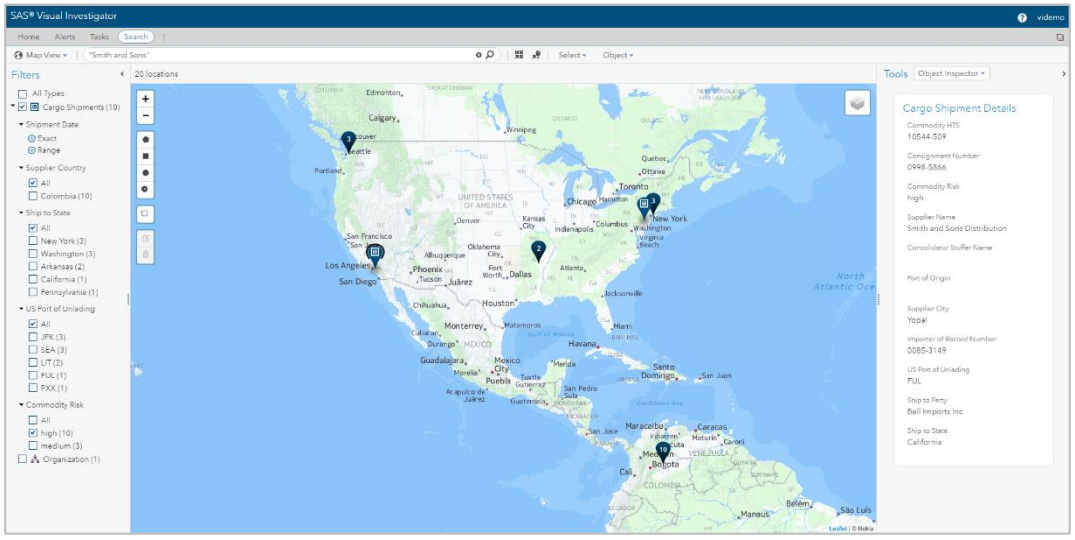



**Figure 4. Exploring Geospatial Data within SAS Visual Investigator**

## ENTITY AND NETWORK ANALYTICS

SAS Visual Investigator uses configurable network building and entity resolution rules to automatically generate social networks based on internal and external data sources. This results in the creation of unique entities that represent real-world objects and their relationships, for example, people, addresses, organizations, telephone numbers, aircraft, and vessels. Entity resolution provides a comprehensive, 360-degree view of all information about a given freight shipment or passenger.

In addition to entity resolution, specific relationships might also be explicitly defined in the data (for example, associations based on a common ID such as a consignment number).

Regardless of how relationships are created, it is the network view, within SAS Visual Investigator that provides analysts with the ability to explore the social networks. Users can add entities to a network view and expand multiple levels of links to bring more related objects onto the canvas and identify commonality across multiple pieces of data.

Often it is the shape of the network that is important, and analysts can adjust the layout as needed. As shown below, data-driven node decorators automatically draw attention to important information within the network (for example, if an individual has previously been denied entry to the country or if a cargo shipment lists restricted or high-risk goods on its manifest). Having this information proactively presented on the network chart means that the analyst doesn't have to open each individual entity to find the most pertinent information.

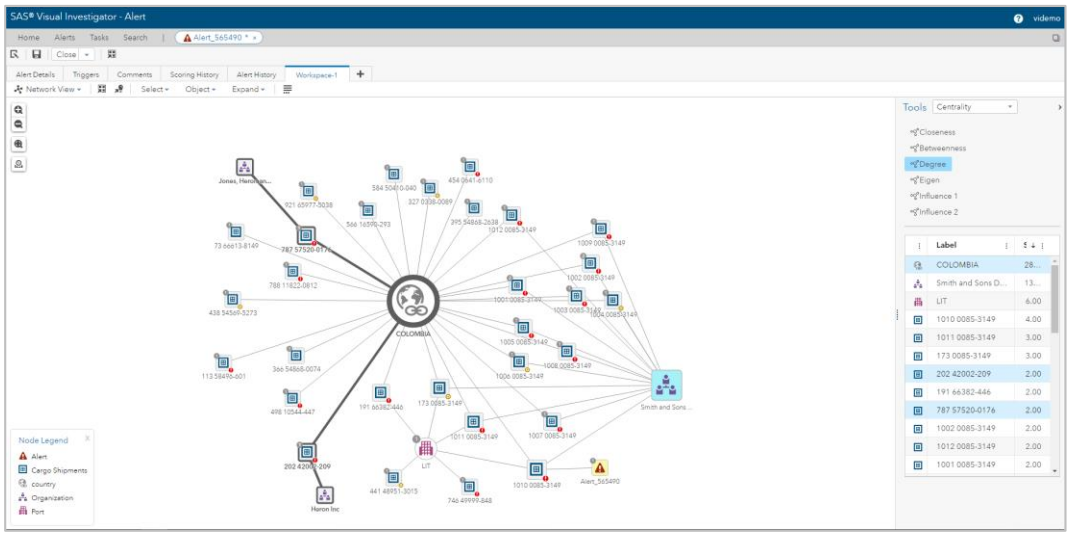



**Figure 5. Exploring Social Networks within SAS Visual Investigator**

Analysts can apply network analytics to their network view in the form of centrality measures to highlight entities of influence or entities most closely linked, or to highlight the shortest path between two nodes in a network.

To help explain a pattern or area of interest within the network, analysts can annotate the view by doing the following:

- changing the node color, icon, and size of nodes
- altering the style of the line used to represent the links
- adding custom nodes and links
- grouping nodes

SAS Visual investigator also provides the ability for analysts to visualize further interactions between two parties by overlaying the social network view with a transactional data flow. The transactions between two

parties might represent financial transactions or other exchanges between two parties such as emails, telephone calls, or the flow of freight between shipping company and the buyer or ship-to party.

## CASE AND INVESTIGATION MANAGEMENT

After the initial alert triage is processed, it might be necessary to conduct further detailed research and investigation on the subject of the alert. Similarly, an analyst or investigator might need to start a case or carry out research based on an ad hoc request or tip from within the organization or received from a partner agency.

During the course of their research, analysts and investigators might follow a different number of leads. However, this is always underpinned by business processes and guidelines. For example, after carrying out initial research, an investigator might need to escalate a specific inquiry to a specialist group of agents who focus on the illegal import of pharmaceutical drugs or counterfeit goods. Being able to package their research, notes, and findings before sending to the appropriate users or teams helps to ensure efficiency while removing the possibility of anything being missed.

SAS Visual Investigator supports case management through the configuration of case and investigation entities that capture details of the work being done and follow a defined workflow that represents the border agencies' own business processes. Workflows might be used to send cases for review, rework, or approval, and to support collaboration between analysts and investigation teams. They can be modeled in a way to ensure border agencies capture institutional experience to guide new or inexperienced agents, analysts, and investigators through the steps required when conducting an investigation or carrying out case work.

Underpinned by workflow, case and investigation entities within SAS Visual Investigator act as work areas in which investigators can do the following:

- gather and explore data in multiple workspace and different visualizations
- add notes and comments
- attach supporting files such as documents, images, and videos
- create Insights reports, as shown below, to record their findings and build an intelligence picture to ratify decisions

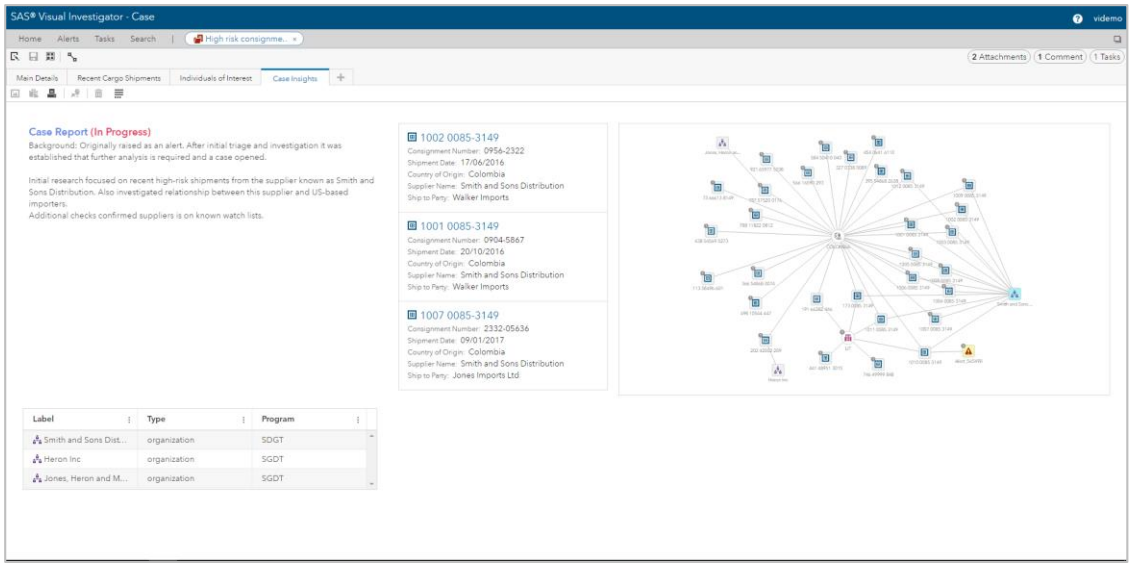



**Figure 6. Managing Cases in SAS Visual Investigator**

The entities and workflows that support case management can be easily updated by administrators on an ongoing basis to allow border agencies to adapt to changes in process and legislation.

## CONCLUSION

Working with silos of information that are difficult for border agents and analysts to access is not suitable for the modern era of border management and international trade. It is imperative that national border agencies empower their analysts through the use of data, risk-based analysis, proactive surveillance and alerting, data exploration, and case and investigation management.

SAS Visual Investigator offers these capabilities through a single-application interface, increasing efficiency, supporting targeting investigations, and enabling agents and analysts to make prompt and informed decisions.

## REFERENCES


International Air Transport Association (IATA). 2015. "Annual Review 2015." http://www.iata.org/about/Documents/iata-annual-review-2015.pdf.

International Air Transport Association (IATA). 2016. "Annual Review 2016." http://www.iata.org/publications/Documents/iata-annual-review-2016.pdf.

United Nations Conference on Trade and Development (UNCTAD). 2016. "Seaborne Trade 2015 Infographic." http://unctadstat.unctad.org/EN/Infographics.html#&gid=2016&pid=Seaborne%20trade%20in%202015.

U.S. Customs and Border Protection. 2016. "Strategic Plan: Vision and Strategy 2020." https://www.cbp.gov/sites/default/files/documents/CBP-Vision-Strategy-2020.pdf.

U.S. Customs and Border Protection. 2017. "Cargo Security and Examinations." Accessed January 19, 2017. https://www.cbp.gov/border-security/ports-entry/cargo-security.


## CONTACT INFORMATION

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


Susan Trueman
SAS Institute Inc.
480 Argyle Street
Glasgow, G2 8NH, United Kingdom
+44 (0)141 223 9100
susan.trueman@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.

Paper SAS495-2017

# Investigating Connections between Disparate Data Sources with SAS® Visual Investigator

Brooke Fortson and Gordon Robinson, SAS Institute Inc.

## ABSTRACT

In 1993, Erin Brockovich, a legal clerk to Edward L. Masry, began a lengthy manual investigation after discovering a link between elevated clusters of cancer cases in Hinkley, CA, and contaminated water in the same area due to the disposal of chemicals from a utility company. In this session, we combine disparate data sources—cancer cases and chemical spillages—to identify connections between the two data sets using SAS® Visual Investigator. Using the map and network functionalities, we visualize the contaminated areas and their link to cancer clusters. What took Erin Brockovich months and months to investigate, we can do in minutes with SAS® Visual Investigator.

## INTRODUCTION

When Erin Brockovich began organizing papers for a pro bono real estate case, she didn't realize that she'd spend countless hours in her law firm connecting the dots between medical records and toxic spills in Hinkley, CA. Brockovich quickly learned how difficult it can be to bring together disparate data sources to expose obscured patterns and reveal hidden connections. After months and months of research, she was able to identify illnesses that correlated with a Chromium 6 toxic spill that had been poisoning Hinkley's water for over 30 years.

In this paper, we tackle an analysis similar to Brockovich's, bringing together data around cancer cases and toxic spills. This analysis helps us to identify cancer clusters in a geographical location and see whether there is any correlation of those clusters to contaminated areas. We create alerts to notify an investigator when there is a possible cancer cluster or new cases within an existing cluster, so that we can take action quickly and confidently.

Here are some things to consider:

- What if the contamination is at a workplace? There might not be a geographical cluster based on home address.
- Cancer could take quite some time to present itself. What if the residents have moved away?

These are realistic scenarios that would be nearly impossible for a human to detect, and they illustrate why it took Erin Brockovich so long to complete her research.

With SAS Visual Investigator configured to look at this data, the movie *Erin Brockovich* would have been a 90-second trailer.

## CONNECTING THE DOTS / MAKING ANALYTICS ACTIONABLE

"It is impossible to connect the dots looking forward but very, very clear looking backwards." – Steve Jobs

While Jobs' observation might be true for humans, SAS Visual Investigator is able to connect the dots in real time using scenarios. In today's world there is so much data. How do you know what is important and what is just noise? What do you prioritize and what can be discarded? When you import millions of rows of data, most of it is going to be meaningless, unimportant information. But what if there's a small but important data point that could get overlooked? Or what if there's a nearly undetectable pattern?

When thinking about SAS Visual Investigator, I always think about the evidence boards in crime shows, where investigators are trying to link together evidence and make connections between people and places. The wall is always covered in maps, timelines, and mug shots, all connected by strings and other evidence. Not only is this relying on investigators to manually detect patterns and connections, but it's also a singular snapshot, only accessible by the people in the room. For Erin Brockovich, her evidence

board consisted of file folders and troves of medical charts, along with thousands of research hours. But with SAS Visual Investigator, you have a virtual environment that enables an investigator to interact with the data, document findings, and share insights with colleagues, making investigations efficient and collaborative.

## SOURCE DATA

There are two main data sources that we are pulling into SAS Visual Investigator.

First, we want to be able to pull in details of the cancer incidents. This would include lots of data points about the individuals affected, including the following:

- age.
- address history. This is important in case the individual has moved, given the latency involved in developing cancer.
- sex.
- ethnicity.
- employment history.
- details of the cancer.
    - type
    - location
    - date of diagnosis

The second data source that we pull in concerns toxic spills that occurred between 1987 and 2015 (https://www.epa.gov/toxics-release-inventory-tri-program). This data set includes details of the following:

- chemical involved
- date of the spillage
- amount of the chemical spilled
- whether the chemical is a carcinogen
- facility at which the spillage occurred
- parent company for the facility

## USING THE DATA WITHIN SAS VISUAL INVESTIGATOR

SAS Visual Investigator provides the capabilities to bring together multiple sources, generate relationships between the data, and provide a single tool for analysts to use to perform their investigations.

The aim of the application is to enable users to leave their data where it is and use existing data sources. That said, in many cases there is some data preparation work that needs to be performed for effective investigations. For example, the address and employment histories should be standardized prior to loading them into SAS Visual Investigator. SAS provides fantastic tooling to perform this standardization.

## CLUSTERING THE CANCER CASES

Prior to loading the cancer cases into SAS Visual Investigator, we want to first identify any geographical clusters within the incidents. There are multiple addresses for each of the cancer incidents, as the individuals involved will likely have resided at multiple addresses throughout their lifetime. It is important to be able to include this movement in the analysis to help us identify geographical clusters.

As part of a data preparation stage, we want to add a column to addresses related to the cancer cases. This column will store a key for any identified clusters. The addresses have all been geocoded, meaning that we can look for clusters based on this information.

Table 1 shows an example of what the resulting table would look like.

| Latitude | Longitude | Cluster |
|---|---|---|
| 35.77966 | -78.5192 | 0 |
| 35.90939 | -78.7898 | 0 |
| 35.91514 | -78.6584 | 1 |
| 35.94685 | -78.7449 | 0 |
| 35.86421 | -78.4804 | 1 |
| 35.79364 | -78.7205 | 1 |
| 35.77286 | -78.7394 | 2 |
| 35.87001 | -78.5129 | 0 |
| 35.76058 | -78.541 | 2 |

**Table 1. Sample Data after Clustering**

We also want to generate some statistics around the cluster. For each cluster, we want to know this information:

- the total number of incidents within each cluster
- the number of incidents of each type of cancer

## IDENTIFYING ENTITY TYPES WITHIN THE DATA

SAS Visual Investigator uses the concept of entities in relation to the data. Data models are based on multiple different entity types. For example, a banking data model is likely based upon these entity types:

- Personal Customers
- Business Customers
- Accounts
- Transactions

Each entity might have multiple tables of associated data. For example, a customer might have multiple addresses, with these being stored in a separate table. This type of relationship is described as a child entity. For example, the Personal Customer entity type might have an Address child entity type.

It's likely there are also relationships between the entities. These take the form of either foreign key relationships within the data or of bridge tables.

When looking at data to integrate into SAS Visual Investigator, some thought has to be given to the entities that the analyst would expect to see and what will lead to most efficient investigations.

The toxic spillage data that we have pulled from the web has an entry for every spillage that has occurred in the US between 1987 and 2015. (There are around 2.5 million spillages recorded in the data.) Each row identifies the following information:

- the chemical involved
- whether the chemical is a known carcinogen
- the amount of the chemical released
- the facility at which the incident occurred
- the parent company associated with the facility

We could present this within SAS Visual Investigator as a Spillage entity type. This wouldn't be particularly useful to the analyst, though.

It would be more interesting to present Company and Facility entities within SAS Visual Investigator and link these together. The Facility entity type would have a child entity called Spills. This would allow the analyst to easily see all of the spillages that have occurred within that facility.

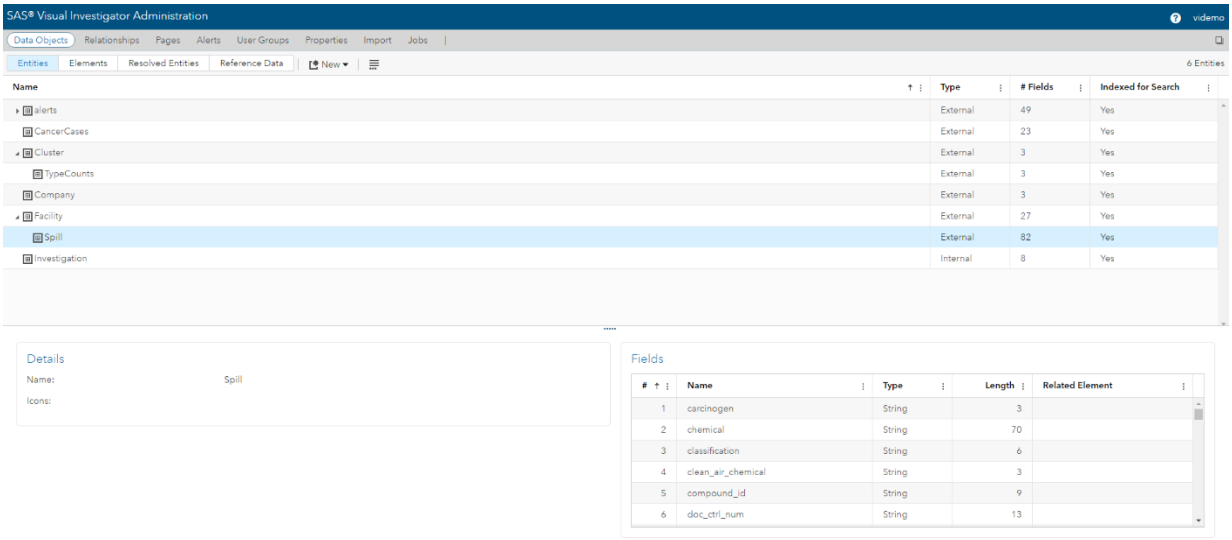
Once the spillage data has been loaded into Postgres, facility and company tables can easily be generated by using some SQL Select Distinct commands. To do this, we need to make sure that we generate the data to be able to link the companies, facilities, and spills together.

Given the data that we have for investigating cancer incidents, we have the following entity types.

- Cancer Incidence
- Cancer Cluster
- Company
- Facility

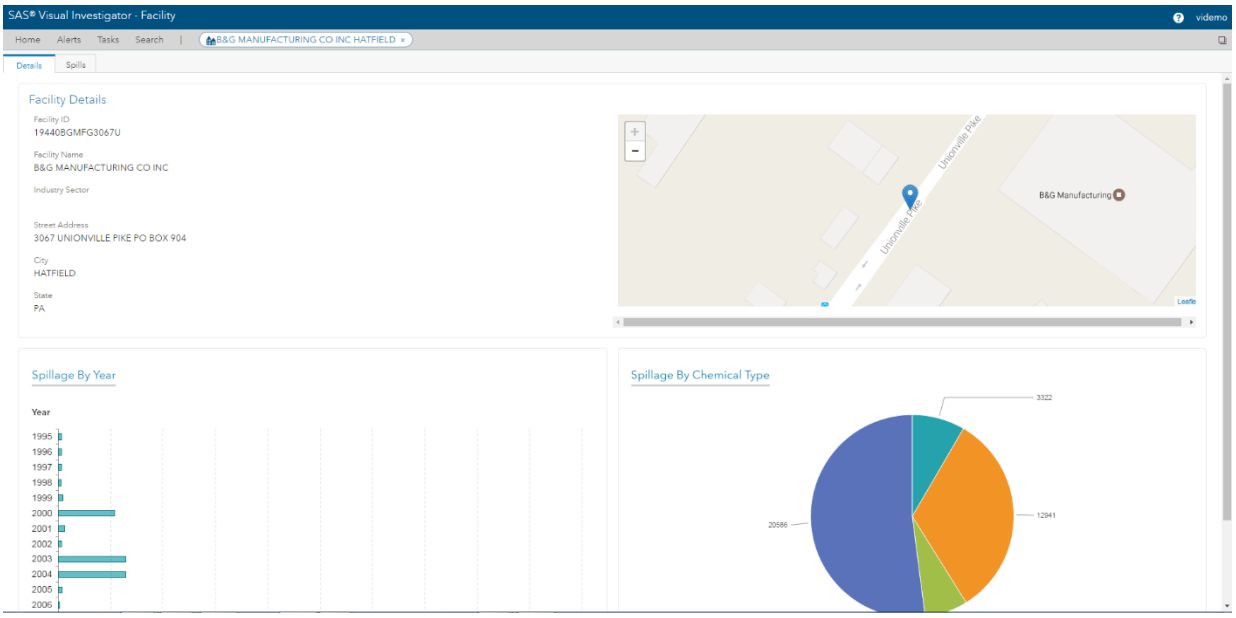
Figure 1 shows a screenshot of the Administration module of SAS Visual Investigator. This tool enables the system administrators to point at data sources and decide which items to pull in.

The administrators also have the ability to define how these entities are related to each other (for example, defining how the Company entities are related to the Facility entities).




**Display 1. Administering Entity Types within SAS Visual Investigator**

One of the key strengths of SAS Visual Investigator is that it enables administrators to define how the entities are presented to the user. The page builder component enables the administrator to drag and drop controls onto a page to define what the analyst would see. Display 2 shows an example of the view that could be defined for a facility.




**Display 2. A Facility Viewed within SAS Visual Investigator**

This view of the facility allows us to easily see details such as the following:

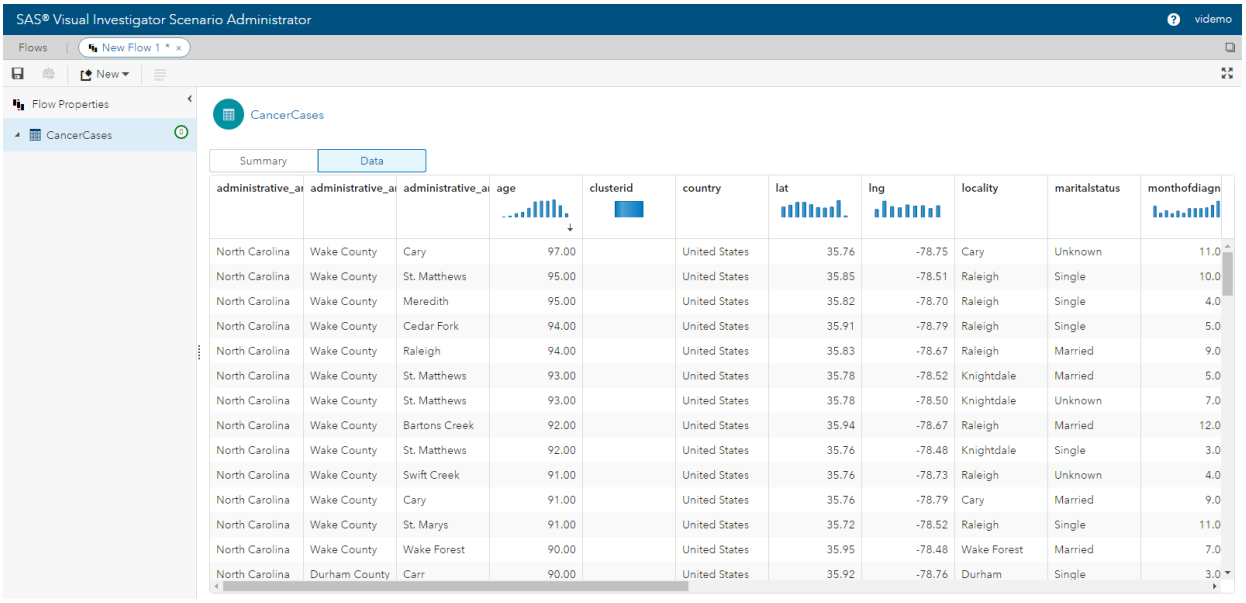
- the name of the facility
- the location, plotted on a map
- the types of chemicals that have been spilt at the facility
- a chart showing the spillages per year

## ALERTING ON CANCER INCIDENTS

The Scenario Administrator functionality of SAS Visual Investigator enables an administrator to define multiple scenarios that can be used to find cases that need to be investigated by an analyst.

Scenario Administrator allows for the following types of scenarios to be created:

- Grouping Scenario – This scenario enables us to group entries in the associated table by either an associated entity type or by another column within the data. We can then look at attributes of the cluster to determine whether alerts should be created. For example, we might want to look at the minimum and maximum age at the time of diagnosis within a cancer cluster.
- Record-Level Scenario – This scenario allows for rules to be run on individual entries within a table. For example, we might want to look at individual cancer incidents where the type of cancer is uncommon for the age of the individual at diagnosis.
- DATA Step Scenario – This scenario allows for DATA step code to be run for more complex scenarios that can't be achieved by the grouping or record-level scenarios.




**Display 3. Scenario Administrator within SAS Visual Investigator**

Each alert generated relates back to one of the entities that we have defined within SAS Visual Investigator. For example, we might wish to alert on the following entities:

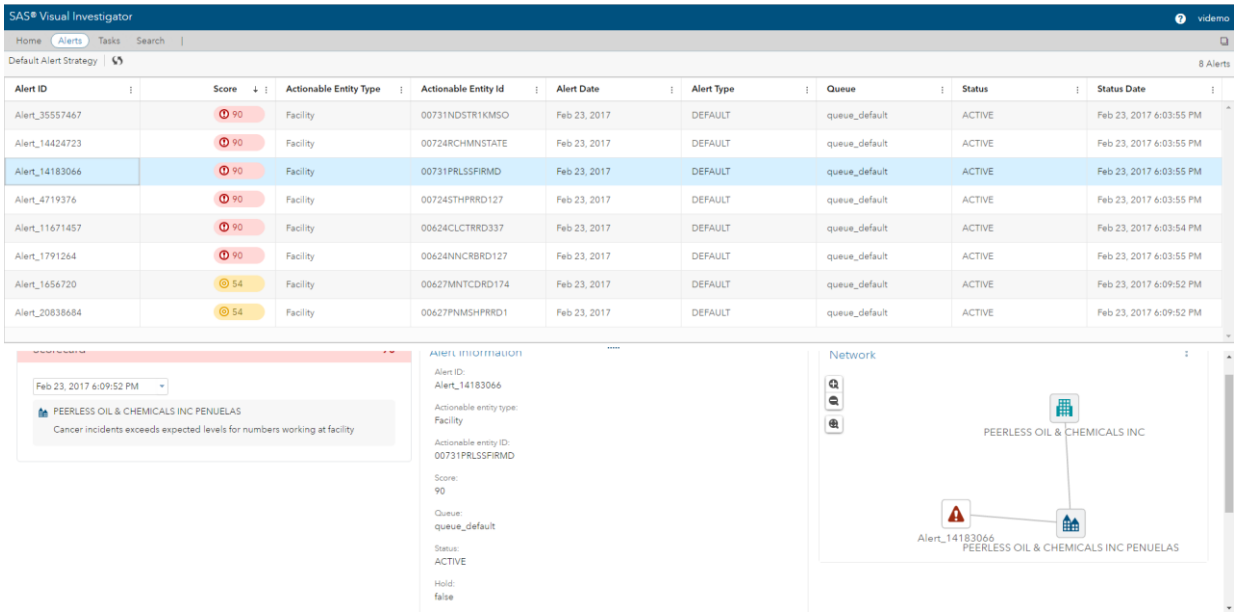
- a Cluster if we find that there are multiple incidences of the same cancer type within it
- a Facility if we find too many cancer cases from employees who worked there
- a Company if we find a higher than normal count at their various facilities

## INVESTIGATING THE ALERTED INCIDENTS

Once alerts have been generated within SAS Visual Investigator, they can be routed to an analyst for investigation. This is achieved through the use of strategies and queues.

Strategies allow for the grouping of alerts to a defined business problem. Within a strategy, there might be multiple queues. Queues are used to control the priority of the alerts within a strategy and how these are dealt with by the analysts.

Display 4 shows the triage page provided to analysts to enable them to triage any alerts that have been generated. From this page, an analyst can view all of the alerts. Double-clicking on a row within the grid opens that alert and provides the analyst with more information about it.

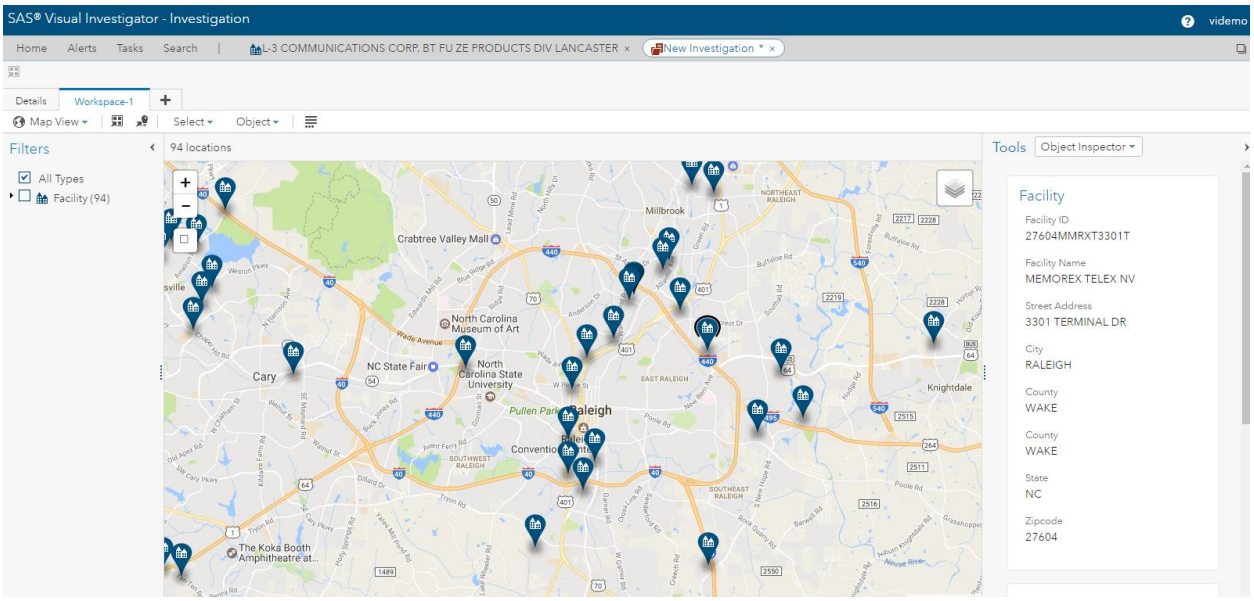



**Display 4. Alert Triage Page within SAS Visual Investigator**

Within an alert, an analyst can start to explore the data within a tool called a workspace. A workspace enables an analyst to collate data that is pertinent to their investigation. Within a workspace, an analyst can choose the following views:

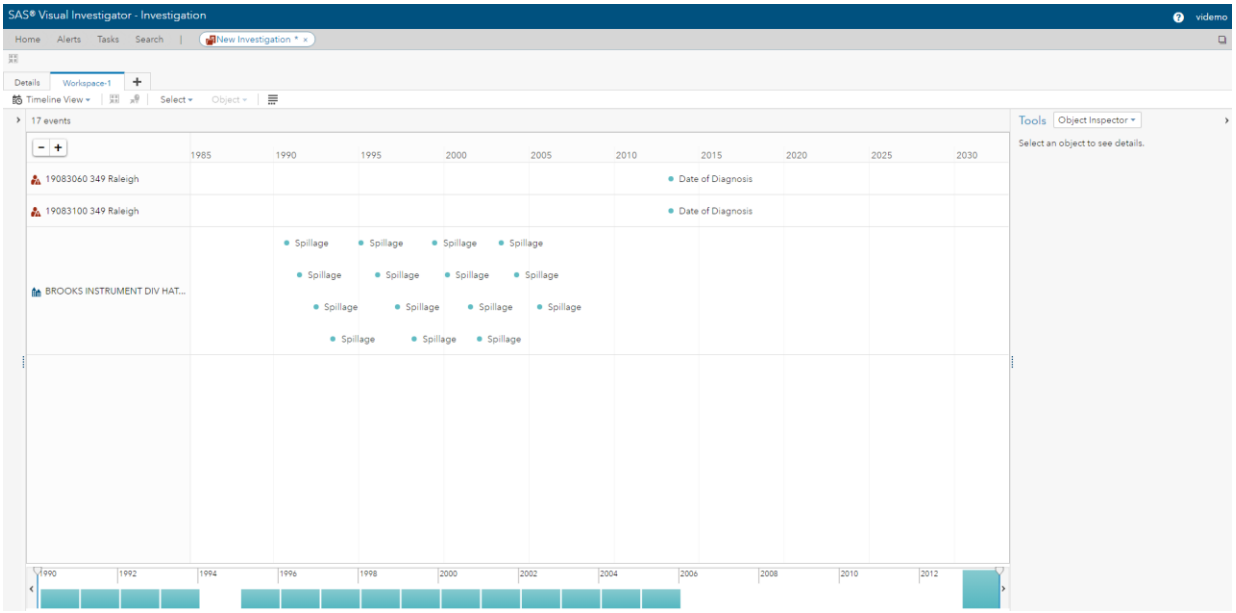
- Detail view shows a list of the items within the workspace.
- Map view plots any location data on a map.
- Timeline view plots any temporal information on a timeline.
- Grid view lists the items in a grid, enabling the analyst to choose the columns to display and to sort on.

If the analyst is investigating an alert based on multiple incidences of cancer of the same type within a geographical area, then they might wish to look for potential causes. They might wish to use the map view, for example, to look for environmental issues that could have influenced the number of incidents. Figure 2 shows the map view of the workspace within SAS Visual Investigator.




**Display 5. Map View within a Workspace**

If the analyst is investigating incidents of cancer within a particular work place, they might wish to use the timeline capability of the workspace to look at when the employees worked at the facility in relation to any toxic spillages that might have occurred. Figure 3 shows the timeline functionality of the workspace.

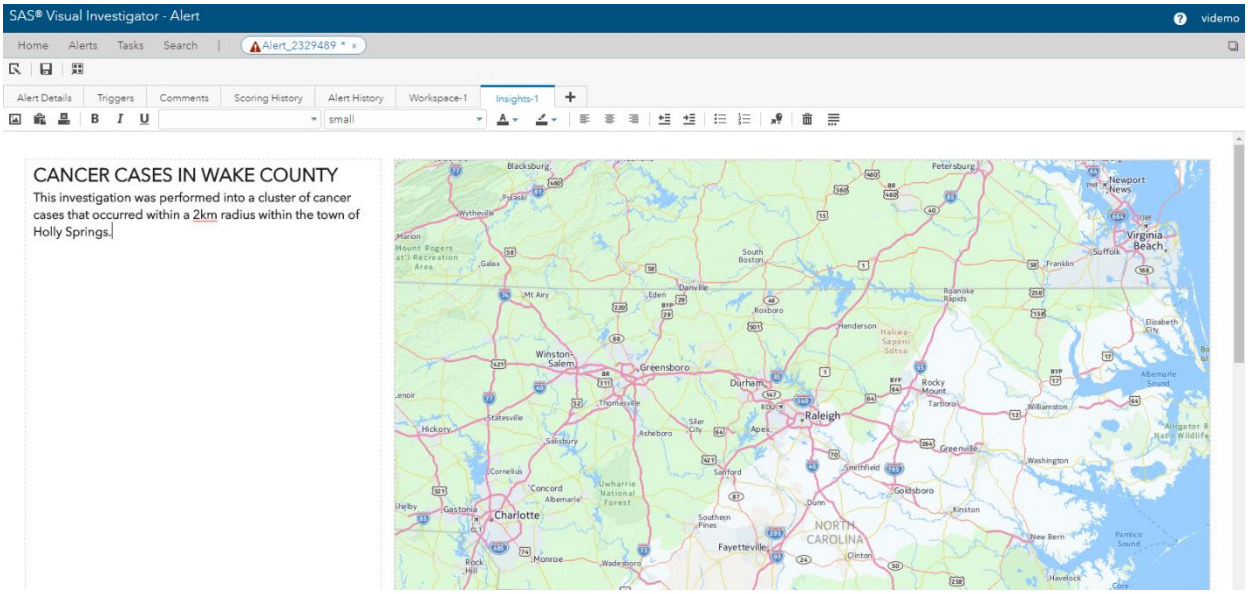



**Display 6. Timeline Feature of the Workspace within SAS Visual Investigator**

## DOCUMENTING AN INVESTIGATION

The workspace provides an analyst with the tools that they need to perform their investigation, whereas an insight enables them to document their findings. Insights are stored within the alert or case they are associated with and are searchable within the application. This means that the results of investigations are searchable and can be used to aid future investigations.

Insights enable the analyst to snapshot copies of the visualizations that they have access to within the workspace. They can then add their notes and any images, such as a picture taken at the site of a chemical spill. Figure 4 shows an insight within the context of an alert.




**Display 7. An Insight Documenting an Investigation**

## CONCLUSION

SAS Visual Investigator provides the capabilities to combine multiple data sources, generate relationships between that data, and provide a single tool for analysts to use to perform their investigations. With the use of strategies and queues, alerts can be generated, prioritized, and routed to the appropriate analyst for more efficient, targeted investigations. And with the use of dynamic workspaces, an analyst can explore data through interactive visualizations and search capabilities, and then easily document and share findings.

The tools available for investigators have changed over the past decade. Investigators are now armed with powerful intelligence analytics to proactively detect anomalies and relationships within their data. Erin Brockovich would have benefitted greatly from SAS Visual Investigator – from combining multiple sources of data, to identifying possible cancer clusters, to documenting her investigation.

Because there is so much fast-moving data in today's world, it is imperative for investigators to keep up. It's no longer acceptable to take months and years to investigate. Investigators need to explore data and conduct investigations with speed and precision.

## REFERENCES

Erin Brockovich. "My Story."Available http://www.brockovich.com/my-story/. Accessed on February 2, 2017.

## RECOMMENDED READING

- *SAS® Visual Investigator 10.2: Administrator's Guide*
- *SAS® Visual Investigator 10.2: User's Guide*

## CONTACT INFORMATION

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

Brooke Fortson
SAS Institute
Brooke.Fortson@sas.com

Gordon Robinson
SAS Institute
Gordon.Robinson@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.

Paper SAS686-2017

# Fighting Crime in Real Time with SAS® Visual Scenario Designer

John Shipway, SAS Institute Inc., Cary, NC

## ABSTRACT

Credit card fraud. Loan fraud. Online banking fraud. Money laundering. Terrorism financing. Identity theft. The strains that modern criminals are placing on financial and government institutions demand new approaches to detecting and fighting crime. Traditional methods of analyzing large data sets on a periodic, batch basis are no longer sufficient. SAS® Event Stream Processing provides a framework and run-time architecture for building and deploying analytical models that run continuously on streams of incoming data, which can come from virtually any source: message queues, databases, files, TCP/IP sockets, and so on. SAS® Visual Scenario Designer is a powerful tool for developing, testing, and deploying aggregations, models, and rule sets that run in the SAS® Event Stream Processing Engine. In this paper, we will explore the technology architecture, data flow, tools, and methodologies that are required to build a solution based on SAS Visual Scenario Designer, enabling organizations to fight crime in real time.

## INTRODUCTION

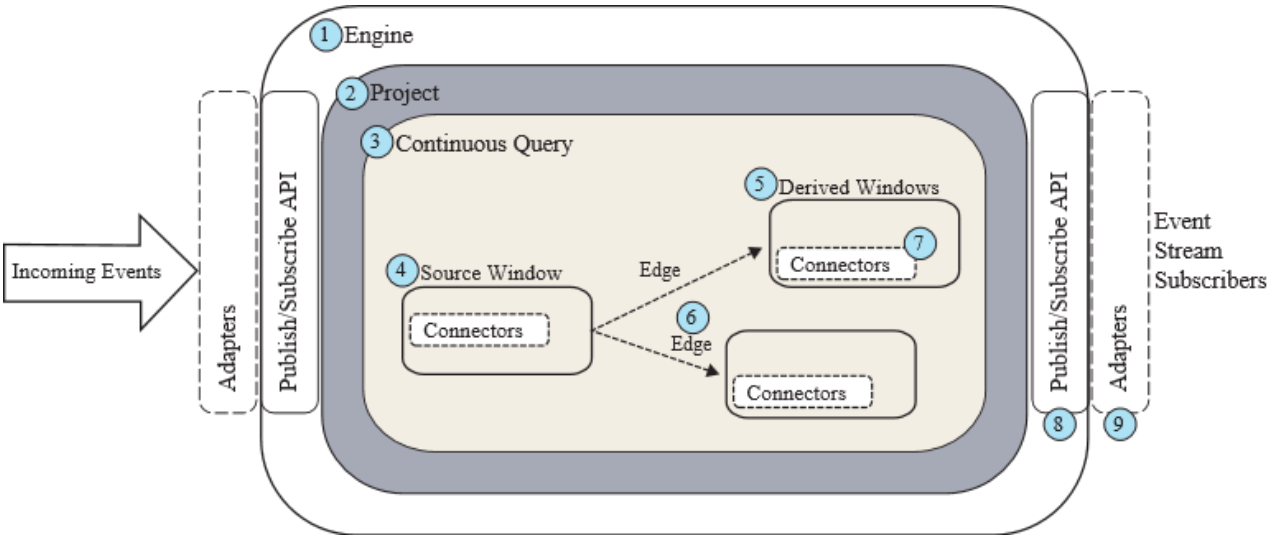
Traditional methods of performing fraud detection have relied on batch runs, periodically analyzing accumulated transactions to identify anomalous behavior.

As fraudsters and criminals have become more sophisticated, and are increasingly committing their crimes electronically, if we wait for the nightly batch run, the criminal might have already changed their identity, their Social Security number, their address, and their phone number by the time we have detected potential fraud.

To address this rapid pace of change within the criminal, SAS is offering real-time analytical capabilities, whereby potential fraud, identity theft, money laundering, and other crimes can be identified *at the time that the transactions occur.* This is the key paradigm shift in the analysis model for crime detection: batch versus real time.

## SAS EVENT STREAM PROCESSING

SAS Event Stream Processing Engine provides this real-time analysis capability. SAS Event Stream Processing provides a generic capability for connecting to external data sources such as message queues, text files, databases, REST interfaces, and TCP/IP connections in order to stream data into continuous queries, as depicted below:




**Display 1. Event Stream Processing, Architectural View**

A continuous query is represented by a directed graph. This graph is a set of connected nodes that follow a direction down one or more parallel paths. Continuous queries are data flows, which are data transformations and analysis of incoming event streams.

Each query has a unique name and begins with one or more source windows.

Source windows are typically connected to one or more derived windows. Derived windows can detect patterns in the data, transform the data, aggregate the data, analyze the data, filter the data, or perform computations based on the data. They can be connected to other derived windows. Ultimately, the results of this event streaming execution can be pushed out to external systems via message queues, text files, databases, REST interfaces, and TCP/IP sockets.

## FRONT RUNNING EXAMPLE

Front running provides a perfect use case for SAS Event Stream Processing in the arena of crime detection. Front running is the practice of a stock broker taking a large order for client, and executing a personal order for themselves *before* the large institutional order is executed. Once the large order is placed, the market adjusts to the newly demonstrated demand, the price of the stock increases, and the broker then sells his shares for a profit.

In order to detect front running, we would develop and deploy a SAS Event Stream Processing model based on a pattern window. A pattern defines events of interest that are associated with each other via logical operators, and which occur within various time-based conditions. For the front running case, the events of interest are as follows:

1) Event #1 (e1): Buy shares of a stock for yourself, Broker = Customer
2) Event #2 (e2) : Buying shares of a stock for a customer, Broker != Customer
3) Event #3 (e3) : Selling your stock for a profit, Sale > Purchase

Events of interest can be combined across events (rows) of the input event stream.

Various operators can be applied to combine events of interest. The “followed by” operator detects whether each event of interest is followed in the event stream by the one following it in the expression list. To detect the front running sequence, we would use the following operator:

```
fby(e1,e2,e3)
```

If we wanted to add a time element that would specify that all the events must occur within a certain time

window, one hour, we could simply add:

```
fby{1 hour}(e1,e2,e3)
```

There are additional operators that could be used within pattern windows. If we wanted to detect the fact that a certain event had not happened within six hour time period, we could use the “does not occur” operator:

```
notoccur{6 hour}(e1)
```

If we wanted to detect the fact that a certain event happened immediately after another event, with no events occurring between the two, we could use the “is” operator.

```
e1 and is(e2)
```

The entire SAS Event Stream Processing XML for front running detection is listed below:

```
<window-pattern name='front_running'>
<schema-
string>id*:string,broker:int32,symbol:string,date1:string,date2:string,date3:
string,id1:int64,id2:int64,id3:int64</schema-string>
<patterns>
<pattern>
<events>
<event name='e1'>((buysellflg=='B') and (broker == buyer) and (s == symbol)
and (b == broker) and (p == price)) </event>
<event name='e2'>((buysellflg=='B') and (broker != buyer) and (s == symbol)
and (b == broker)) </event>
<event name='e3'><![CDATA[buysellflg=='S') and (broker == seller) and (s ==
symbol) and (b == broker) and (p < price))]]></event>
</events>
<logic>fby(e1,e2,e3)</logic>
<output>
<field-selection name='broker' node='e1'/>
<field-selection name='symbol' node='e1'/>
<field-selection name='date' node='e1'/>
<field-selection name='date' node='e2'/>
<field-selection name='date' node='e3'/>
<field-selection name='id' node='e1'/>
<field-selection name='id' node='e2'/>
<field-selection name='id' node='e3'/>
</output>
</pattern>
</patterns>
</window-pattern>
```

## TERRORISM SURVEILLANCE EXAMPLE

A common real-time pattern in identifying potential terrorist activity is in the area of flight ticket purchases. Purchases by cash or Visa gift card raise a flag. Long distance flights from a major metropolitan hub, to countries known for terrorist financing training, and back again in a short time window are suspect. This is a perfect real-time surveillance use case, as often the tickets are initially bought with a return of several days later, outside the window of suspicion, but are then later changed to a flight an hour from the time of purchase, the day after they arrived. If we waited to analyze all of the flight transactions in a nightly batch run, the individual would likely be back to the major metropolitan hub by the time the anomaly was found, armed with whatever they picked up on the trip. By analyzing the transactions *as they occur*, and sending alerts immediately to an investigation system such as SAS® Visual Investigator, authorities could be waiting for the suspects while they are waiting to get on the plane for their return flight home, and thwarting a potential attack. For this surveillance scenario, once again, a pattern window could be deployed that checks for a sequence of events as follows:

1) Event 1 (e1): An individual leaving from a list of identified airports (A), going to list of identified airports (B)
2) Event 2 (e2): The same individual booking a flight from (B) back to (A)
3) These two events occurring within a thirty six hour window.
    a. In the SAS Event Stream Processing model, use: `fby{36 hour} (e1, e2)`

## AN ANALYST INTERFACE IS NEEDED

The contents of these streaming continuous queries, particularly those related to fraud detection and crime prevention, can get quite complicated. There are over a dozen windows types in SAS Event Stream Processing. Building SAS Event Stream Processing models, which are represented by an XML grammar, is much like programming. A visual tool for developing these XML models, SAS Event Stream Processing Studio, has been available, but the user of SAS Event Stream Processing Studio still worked at a very low level, wiring together dozens to hundreds of windows manually. For the fraud or security intelligence analyst, a higher level tool is needed to allow the user to codify and express higher level domain specific “scenarios”, and generate the appropriate SAS Event Stream Processing XML. This higher level tool is SAS Visual Scenario Designer.

# SAS VISUAL SCENARIO DESIGNER – BACKGROUND AND EVOLUTION

SAS Visual Scenario Designer was originally released in 2014 on the SAS 9.4 platform, the SAS® LASR™ Analytic Server, and the SAS Event Stream Processing Engine. The purpose of SAS Visual Scenario Designer was to allow fraud and compliance analysts to explore and visualize their data, perform data-preparation operations, and author “scenarios” that detect anomalous behavior. Scenarios are encapsulated pieces of logic that determine whether a given account, customer, patient, doctor, or other entity represented in a data source is exhibiting behavior that warrants additional investigation. The canonical example of a scenario is “Alert on all the banking customers that have deposited over $10,000 in cash over the last seven days”. Another simple example would be, “Alert on all the physicians that are prescribing a drug or medical device 50% more than their peers in the same specialty.” SAS Visual Scenario Designer scenarios were developed, tested, and simulated exclusively against the SAS LASR Analytic Server, and could ultimately be deployed for production use into a SAS LASR Analytic Server environment or into SAS Event Stream Processing. Batch processing was typically performed in SAS LASR Analytic Server, and real-time/streaming processing was performed in SAS Event Stream Processing. While this provided a unified authoring and simulation environment between batch and real time, there were several differences between them that made testing and simulating against SAS LASR Analytic Server and deploying to SAS Event Stream Processing problematic:

- Time unit differences: SAS LASR Analytic Server supports WEEK/QUARTER/YEAR, SAS Event Stream Processing has no higher grain than DAY.

- Lookback behavior: Default SAS LASR Analytic Server behavior is that a DAY lookback goes from the transaction timestamp to the first tick of the current day. SAS Event Stream Processing behavior is that a DAY lookback goes 24 hours plus one tick from the transaction timestamp.
- Function differences: SAS LASR Analytic Server uses SAS functions. SAS Event Stream Processing uses DataFlux functions. The function signature and behavior of functions of the same category was not always identical.

One other key characteristic of SAS Visual Scenario Designer is that it relied on an overall alert management process outside of the solution. This overall process typically includes alert triaging, case management, case investigation, workflow, and alert dispositioning. When anomalous behavior was detected, and alerting events were generated, some other process or system needed to be developed or integrated with to handle those events.

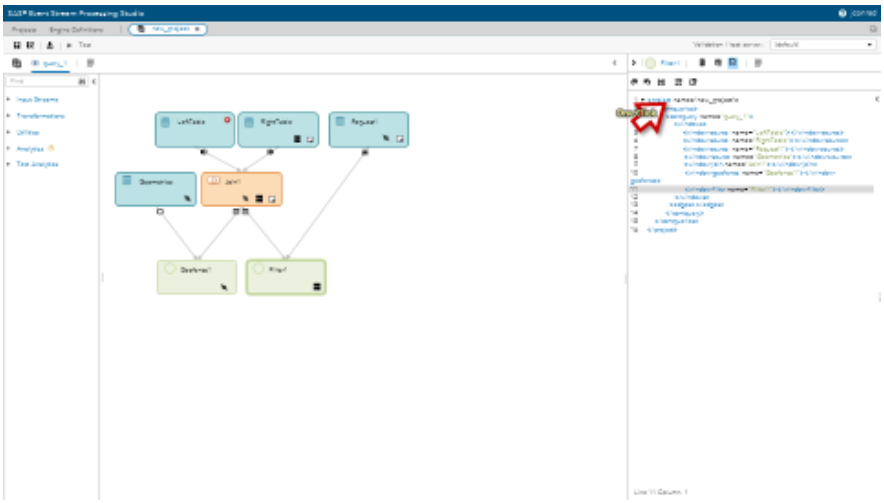
## SAS VISUAL INVESTIGATOR –THE NEW SAS SURVEILLANCE AND INVESTIGATION PLATFORM

As SAS began looking toward to the future with SAS® Viya™ and SAS® Cloud Analytic Services (CAS), a turn-key surveillance and investigation system was envisaged, encompassing real-time surveillance, batch surveillance, data source management, alert triaging, case management, network and relationship analysis, workflow, and alert dispositioning. The 10.1 version of this new surveillance and investigation system, SAS Visual Investigator, was released in September 2016. The 10.2 version was released in March of 2017. A key design goal for SAS Visual Investigator was to provide all the elements of an alert generation and alert management process in one unified system. Another key design goal was to provide open interfaces for sending alerts to SAS Visual Investigator, such that a variety of surveillance approaches could be applied, each sending alerts to SAS Visual Investigator as needed. This open interface into the alerting system of SAS Visual Investigator was achieved via message queues and a well-defined alert document specification. Any system using the appropriate alert JSON format, and publishing alerts to the SAS Visual Investigator alert message queue, could participate in the SAS Visual Investigator surveillance infrastructure.

SAS Visual Scenario Designer underwent a significant re-architecture for the 10.2 SAS Viya release. Rather than try to author and test batch and real-time scenarios with a single interface, the batch functionality of SAS Visual Scenario Designer was re-branded into the Scenario Administrator component of SAS Visual Investigator. The real-time scenario authoring capabilities of SAS Visual Investigator were moved into SAS Event Stream Processing Studio, by way of a newly released extensibility capability, called the SAS Event Stream Processing Studio plug-in interface.

## SAS EVENT STREAM PROCESSING STUDIO AND THE SAS VISUAL SCENARIO DESIGNER SURVEILLANCE PLUG-INS

With the 4.3 release of SAS Event Stream Processing Studio in March 2017, SAS Event Stream Processing Studio enabled other solutions to create higher level real-time constructs. A typical SAS Event Stream Processing Studio model is depicted below:



**Display 2. SAS Event Stream Processing Studio Process Flow Diagram**

Each node in the process flow diagram represents a low-level SAS Event Stream Processing window, source windows, a join window, a geo-fencing window, and a filter window. With the new plug-in, a new process flow diagram node could be created by a solutions team at SAS, by another software company, by a consultant in the field, by virtually anyone. Ultimately, these plug-in nodes generate a set of native SAS Event Stream Processing window primitives, but the end user doesn't need to know or care about the underlying window structure, they simply use the higher level construct, which comes with its own property pane. A surveillance "scenario" node might generate several source windows combined with a join window, a number of aggregation windows, and a filter window to execute the rule set. This plug-in capability allows SAS Event Stream Processing to be extended into virtually any domain, and will create a new plug-in market.

The first plug-in being developed by the SAS Visual Scenario Designer development team is for formatting an incoming stream into the correct JSON structure to be sent to the SAS Visual Investigator alert message queue. The key properties needed to send such an alert message to SAS Visual Investigator are as follows:

1) Entity Type. What are we investigating (CUSTOMER, ACCOUNT, POLICY, PHYSICIAN, PATIENT, and so on). These types come from the list of entities registered in the SAS Visual Investigator Data Hub.
2) Entity ID Field. For the incoming stream, what field represents the unique identifier of the Entity Type specified above.
3) Alert Queue. This is not to be confused with the RabbitMQ message queue. The Alert Queue is a routing construct within the SAS Visual Investigator alerting component.
4) Scenario ID: This is a unique identifier for the scenario that generated this alert. Each SAS Event Stream Processing surveillance model will be assigned a unique identifier so that the alerting module within SAS Visual Investigator knows how the alert was generated.

The SAS Visual Investigator alert plug-in will surface a properties pane with the four properties listed above, and will ultimately generate SAS Event Stream Processing procedure window XML, which creates a JSON string with the appropriate format to be ingested by SAS Visual Investigator. An example of this procedure window XML is below:

```
<window-functional name="Functional_1" pubsub="true" collapse-updates="true">
  <schema>
    <fields>
      <field name="sequence" type="int64" key="true"/>
      <field name="json_format" type="string"/>
    </fields>
  </schema>
  <function-context>
    <functions>
      <function name="json_format"><![CDATA[string('{
    "alertingEvents": [
        {
            "alertingEventId": "',$alertingEvents_alertingEventId,'",
            "actionableEntityType":
"',$alertingEvents_actionableEntityType,'",
            "actionableEntityId": ',$alertingEvents_actionableEntityId,',
            "score": ',$alertingEvents_score,',
            "alertOriginCd": "',$alertingEvents_alertOriginCd,'",
            "alertTypeCd": "',$alertingEvents_alertTypeCd,'",
            "alertTriggerTxt": "',$alertingEvents_alertTriggerTxt,'",
            "scenario_id": ',$alertingEvents_scenario_id,',
            "recQueueId": "',$alertingEvents_recQueueId,'"
        }
    ],
    "scenarioFiredEvents": [
        {
            "scenarioFiredEventId":
"',$scenarioFiredEvents_scenarioFiredEventId,'",
            "alertingEventId": "',$scenarioFiredEvents_alertingEventId,'",
            "displayTypeCd": "',$scenarioFiredEvents_displayTypeCd,'",
            "displayFlg": ',$scenarioFiredEvents_displayFlg,',
            "scenarioOriginCd": "',$scenarioFiredEvents_scenarioOriginCd,'",
            "scenarioDescriptionTxt":
"',$scenarioFiredEvents_scenarioDescriptionTxt,'",
            "scenarioId": "',$scenarioFiredEvents_scenarioId,'"
        }
    ],
    "enrichment": [
        {
            "alertingEventId": "',$enrichment_alertingEventId,'",
            "url": "',$enrichment_url,'",
            "date_published": "',$enrichment_date_published,'",
            "title": "',$enrichment_title,'"
        }
    ]
}')]]></function>
    </functions>
  </function-context>
</window-functional>
```

You can see that simply dragging a SAS Visual Investigator alert node on the SAS Event Stream Processing Studio canvas, and filling out a few properties, is far simpler than hand-rolling the above XML. The SAS Event Stream Processing Studio plug-in interface is a game changer for event stream processing at SAS.

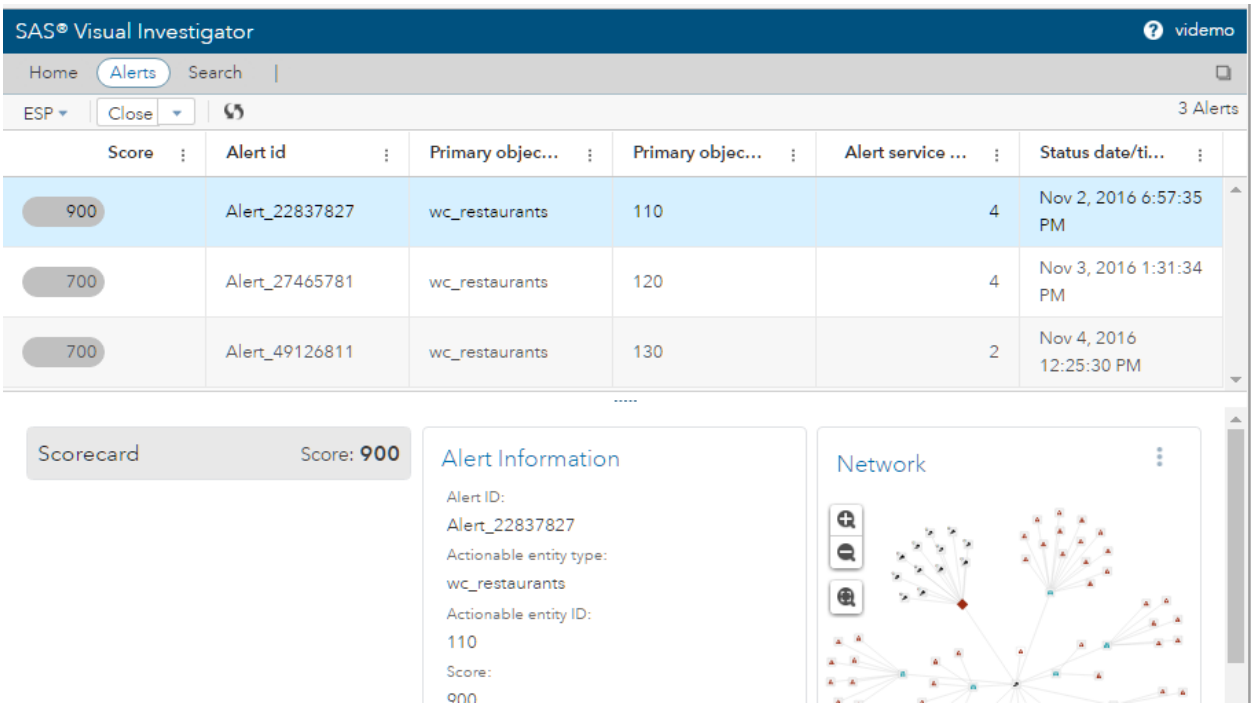
## RABBIT MQ CONNECTOR

In order for the message created above to be sent to SAS Visual Investigator, it must be placed on the SAS Visual Investigator alert message queue. A SAS Event Stream Processing connector node can be configured as shown in the example below:

```
<connectors>
  <connector name="New_Connector_1" class="fs">
    <properties>
      <property name="type"><![CDATA[sub]]></property>
      <property name="snapshot"><![CDATA[true]]></property>
      <property name="fsname"><![CDATA[output.json]]></property>
      <property name="fstype"><![CDATA[json]]></property>
    </properties>
  </connector>
  <connector name="rmq_sub" class="rmq">
    <properties>
      <property name="type"><![CDATA[sub]]></property>
      <property name="rmquserid"><![CDATA[guest]]></property>
      <property name="rmqpassword"><![CDATA[guest]]></property>
      <property name="rmqhost"><![CDATA[10.38.13.246]]></property>
      <property name="rmqport"><![CDATA[5672]]></property>
      <property name="rmqexchange"><![CDATA[svi.tdc.exchange]]></property>
      <property name="rmqtopic"><![CDATA[svi.tdc.ae.q]]></property>
      <property name="rmqtype"><![CDATA[json_format]]></property>
      <property name="numbufferedmsgs"><![CDATA[50000]]></property>
      <property name="urlhostport"><![CDATA[esp-base:39908]]></property>
      <property name="buspersistence"><![CDATA[true]]></property>
      <property name="snapshot"><![CDATA[false]]></property>
      <property name="collapse"><![CDATA[true]]></property>
    </properties>
  </connector>
</connectors>
```

## THE END GAME: ALERTS IN SAS VISUAL INVESTIGATOR

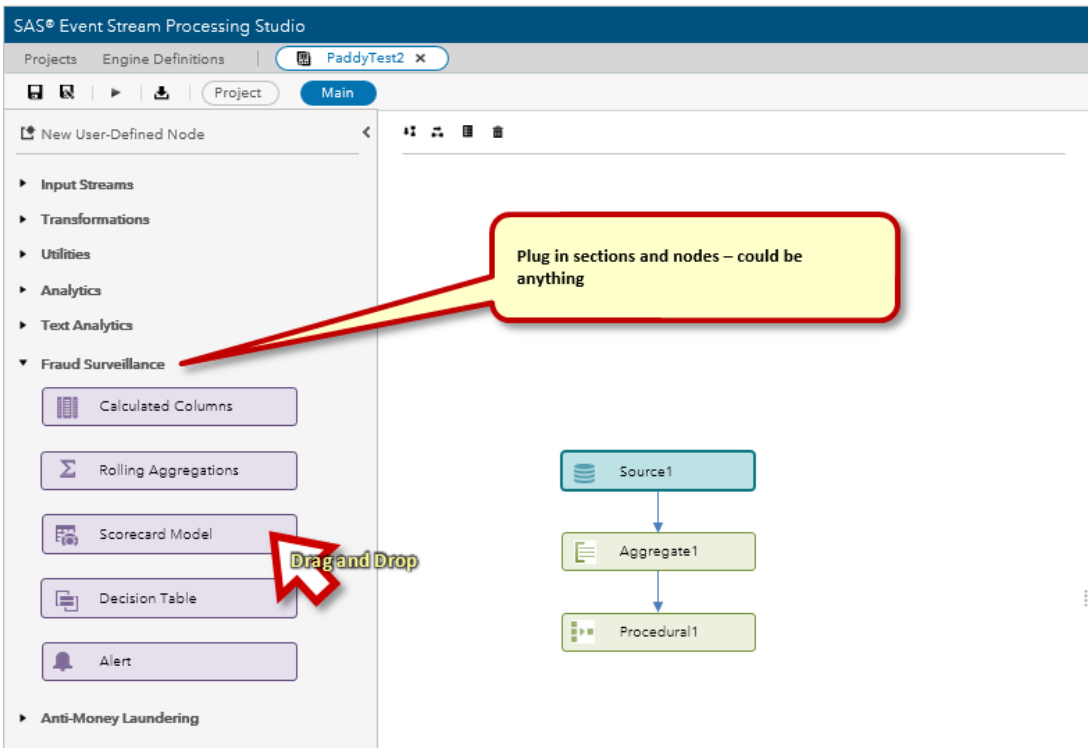
By using a Rabbit MQ Connector in our surveillance model, we are able to send alerts to SAS Visual Investigator from SAS Event Stream Processing. Below is a depiction of the incoming alert queue within SAS Visual Investigator. From this point, an investigator can raise/lower priority of the alert, can create a case for the alert, can initiate a workflow for the alert, can view all the entities related to this alert (customers, accounts, policies), and entities related to those entities, and so on, in a network visualization.




**Display 3. Main Alert View in SAS Visual Investigator**

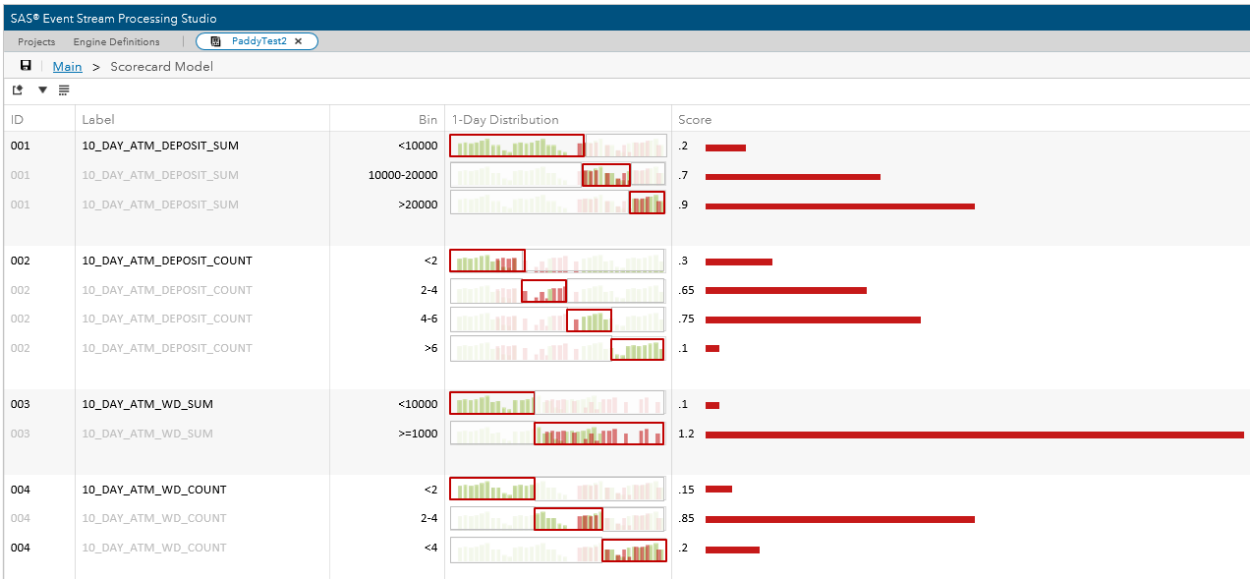
## FUTURE PLUG INS

SAS is actively designing and developing a number of surveillance plug-ins for SAS Event Stream Processing Studio. We envision that various solution teams will create entire plug-in sections that can be surfaced through the process flow diagram pallet, as depicted below:




**Display 4. SAS Event Stream Processing Studio Plug-in Extensions List**

Two powerful constructs used in surveillance scenarios are decision tables and scorecard models. Decision tables allow for a highly flexible if, then, else, else, else heuristic rule creation. Scorecard models allow multiple aggregations to be considered jointly, with fine tuning of lookback time periods (1 DAY, 10 SECONDS, 5 MINUTES), aggregation functions (SUM, COUNT, MEAN, STANDARD DEVIATION, and so on). A design for a forthcoming scorecard model plug-in is depicted below:




**Display 5. Scorecard Model Plug-in**

## ADVANCED ANALYTICS IN REAL TIME

The preceding examples use relatively straightforward event sequence analysis, aggregations, and rule filtering. The SAS Event Stream Processing Engine can handle far more advanced analytics in order to surveil incoming data streams. Analytical models built with SAS® Enterprise Miner™ or SAS® Visual Statistics, and exported as DATA step or DS2 code can be used in SAS Event Stream Processing via a procedure window. Neural networks, decision trees, logistic regression, and linear regression models can all be resolved to DATA step code that SAS Event Stream Processing can use via a procedure window. Further, analytical models developed in Python can also be natively embedded into SAS Event Stream Processing models. Integration with SAS® Model Manager for model tuning, model comparison, and model deployment is in the works. The full power of the SAS 40+ year history in advanced analytics can be used in streaming, real-time models.

## CONCLUSION

Batch analysis of data and periodic alert generation has been the standard for financial crimes and security intelligence investigations for a number of years. With the proliferation of the Internet, online banking, identity theft, hacking, phishing, and electronic crime, the modern criminal has often already erased their footprints, destroyed their fraudulent identity, and electronically moved on to their next identity by the time that the batch analysis has completed and human investigators have been alerted to suspicious behavior. In the age of the digital criminal, real-time analysis and alert generation, occurring at the time that events are occurring, is proving to be a far more effective method of crime detection and prevention. SAS has an event stream processing engine for such real-time analysis detection. With the advent of the SAS Event Stream Processing Studio plug-in interface and the forthcoming SAS Visual Scenario Designer surveillance plug-ins, SAS is empowering organizations to take surveillance authoring to the next level and to fight crime in real time.

## REFERENCES

SAS Institute Inc 2016. Help for SAS Event Stream Processing 4.2. Cary, NC: SAS Institute Inc. http://go.documentation.sas.com/?cdcId=espcdc&cdcVersion=4.2&docsetId=espov&docsetTarget=home.htm&locale=en

SAS Institute Inc. SAS Event Stream Processing 4.2 Training Manual. Cary, NC: SAS Institute Inc.

## ACKNOWLEDGMENTS

Paddy Fahey, Senior User Experience Designer, SAS Institute Inc. for his help in understanding Pattern Windows, and his design content on plug-in sections within SAS Event Stream Studio and the Scorecard model plug-in.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

John Shipway
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
john.shipway@sas.com
http://www.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.

Paper SAS2603-2015

# Addressing AML Regulatory Pressures by Creating Customer Risk Rating Models with Ordinal Logistic Regression

Edwin Rivera, Jim West, and Carl Suplee, SAS Institute Inc.

## ABSTRACT

With increasing regulatory emphasis on using more scientific statistical processes and procedures in the Bank Secrecy Act/Anti-Money Laundering (BSA/AML) compliance space, financial institutions are being pressured to replace their heuristic, rule-based customer risk rating models with well-established, academically supported, statistically based models. As part of their customer-enhanced due diligence, firms are expected to both rate and monitor every customer for the overall risk that the customer poses. Firms with ineffective customer risk rating models can face regulatory enforcement actions such as matters requiring attention (MRAs); the Office of the Comptroller of the Currency (OCC) can issue consent orders for federally chartered banks; and the Federal Deposit Insurance Corporation (FDIC) can take similar actions against state-chartered banks.

Although there is a reasonable amount of information available that discusses the use of statistically based models and adherence to the OCC bulletin *Supervisory Guidance on Model Risk Management* (OCC 2011-12), there is only limited material about the specific statistical techniques that financial institutions can use to rate customer risk. This paper discusses some of these techniques; compares heuristic, rule-based models and statistically based models; and suggests ordinal logistic regression as an effective statistical modeling technique for assessing customer BSA/AML compliance risk. In discussing the ordinal logistic regression model, the paper addresses data quality and the selection of customer risk attributes, as well as the importance of following the OCC's key concepts for developing and managing an effective model risk management framework. Many statistical models can be used to assign customer risk, but logistic regression, and in this case ordinal logistic regression, is a fairly common and robust statistical method of assigning customers to ordered classifications (such as Low, Medium, High-Low, High-Medium, and High-High risk). Using ordinal logistic regression, a financial institution can create a customer risk rating model that is effective in assigning risk, justifiable to regulators, and relatively easy to update, validate, and maintain.

## INTRODUCTION

Assessing customer risk is an essential component of a comprehensive Bank Secrecy Act/Anti-Money Laundering (BSA/AML) monitoring program, because a financial firm's ability to identify customers that pose a higher risk for money laundering and terrorist financing is key to implementing effective customer due diligence (CDD) policies, procedures, and processes. As the Federal Financial Institutions Examination Council's (FFIEC) *BSA/AML Examination Manual* clearly explains, the concept of CDD begins with verifying each customer's identity and assessing both the specific risk and the risk level associated with that customer, as well as putting processes in place to ensure the additional scrutiny that higher-risk customers require.

As part of CDD, firms are expected to both rate and monitor their customers for the overall risk that they pose. Because of the increased regulatory emphasis in the BSA/AML community on using statistical processes and procedures to ensure compliance, banks and other financial institutions face pressure from regulators to stop using heuristic, rule-based customer risk rating processes and instead switch to academically supported, statistically based models. As these firms consider their model options, they must also take into account the Office of the Comptroller of the Currency's *Supervisory Guidance on Model Risk Managemen*t (OCC 2011-12), which outlines the modeling development process, validation requirements, and governance framework that must accompany the model.

Customer risk rating models are systems, algorithms, or processes that firms use either to assign customers to various risk groups or to score the customer's relative risk based on that customer's inherent characteristics, expected transactional behaviors, or overall status at the firm. These models usually fall into either of two primary types: heuristic, rule-based models and statistically based models. As firms look to improve their current customer risk rating models or to implement models where they currently don't exist, questions often arise, such as:

- What are the advantages and disadvantages of using one type of model versus the other?
- Why the regulatory push toward using statistically based models?
- What type of statistically based models should the firm implement?
- What attributes should the firm consider when developing the model?

## HEURISTIC, RULE-BASED MODELS VERSUS STATISTICALLY BASED MODELS

Traditionally, financial firms have created customer risk rating models by focusing on rating customers' risk in several distinct areas, often with multiple variables in a single area. Here are some of the variable types that a customer risk rating model can include:

- *Customer Relationship* (personal, business, commercial, etc.)
- *Geography* (country of residence, business location, High Intensity Financial Crime Areas (HIFCA), High Intensity Drug Trafficking Areas (HIDTA), port or border cities, etc.)
- *Account Features* (remote deposit capture (RDC), correspondent banking, online banking, custodial accounts, etc.)
- *High-Risk Customer* (non-resident alien (NRA), politically exposed person (PEP), money service business (MSB), employee, etc.)
- *Alert / Filing History* (manual alerts created, system-generated alerts, Cash Transaction Reports (CTRs), Suspicious Activity Reports (SARs), etc.)
- *Expected Product Usage* (wires—domestic or foreign, cash, Automated Clearing House (ACH), check, etc.)
- *Expected Transactional Activity* (aggregate dollar amount of activity expected)

Although both heuristic, rule-based models and statistically based models consider the same basic set of customer data attributes, the underlying methodology that each type of model uses to weight the variables and score the customers differs—often significantly.

### HEURISTIC, RULE-BASED MODELS

A heuristic, rule-based model is simply an analytical formula used to assign a score based on one or more variables, or attributes, that the firm deems important. Often these models are created using all available variables, because the relative importance of each individual variable is generally unknown and thus variable selection is quite difficult. As Figure 1 shows, these models are often parameterized so that you can adjust the scores and weights assigned to each component of the model.

| Variable Type | Attributes | Logic Description |
|---|---|---|
| Customer Relationship | Customer Type | + If the customer is “Personal,” then the score is 35.<br>+ If the customer is “Commercial,” then the score is 20. |
| High Risk Customer | Money Services Business (MSB) | + If the customer is “MSB,” then the score is 80. |
| High Risk Customer | Politically Exposed Person (PEP) | + If the customer is “PEP,” then the score is 80. |
| Alert / Filing History | Suspicious Activity Reports (SARs) | + If the SAR count equals 1, then the score is 45.<br>+ If the SAR count is greater than 1, then the score is 60. |
| Expected Transaction Activity | Total Aggregated Transactions | + If the monthly transaction volume is less than $50K, then the score is 40.<br>+ If the monthly transaction volume is greater than or equal to $50K, then the score is 60. |

**Figure 1. Example of a Heuristic, Rule-Based Customer Risk Rating Model**

Each individual customer’s scores are then aggregated, and the firm assigns a risk category based on the customer’s aggregate score. Usually this type of model is constructed based on subject matter expert judgment or knowledge about the underlying process rather than formal analytical analysis. Because there is no specific underlying methodology or model design that firms must follow, they have an endless supply of model design and scoring options to choose from. However, this is also the weakness of these types of models, because the lack of a single statistical framework means there is no established statistical methodology for setting parameters or selecting variables to include in the model. Even after choosing a general modeling framework, firms must make numerous iterative adjustments to determine the combination of parameter settings that maximize the model’s fit to the target variable. In addition, there is no effective way to know whether the chosen parameter set is really the optimal set of values. This makes justifying these types of models to regulators more and more difficult.

Although heuristic, rule-based models were once the norm within the AML community because of their simplicity and ease of development, they are quickly being replaced by more scientific modeling approaches that can stand up to regulators’ scrutiny and that allow for a methodical approach to parameter setting and model validation.

## STATISTICALLY BASED MODELS

Statistically based models are founded on well-established statistical methodologies and approaches that have been vetted, reviewed, and published in academic journals. Most statistically based models that financial firms use for customer risk rating are predictive models, such as linear regression, binary or ordinal logistic regression, decision trees (all types), and neural networks. The particular application and risk rating objectives determine the actual model that the firm selects. However, for customer risk rating, either binary or ordinal logistic regression models are currently the most common.

Unlike heuristic, rule-based models, statistically based models require that certain assumptions be met so that the modeling framework can be accurately tested and assessed to an acceptable degree of statistical confidence. A common goal in creating statistical models is to develop the simplest model—the one with the fewest variables—that is needed to make an accurate prediction. To select those variables, the firm must use a robust statistical framework based on widely accepted modeling approaches that maximize the likelihood that the target is estimated as accurately as possible. In addition, the firm can use standard approaches for assessing each variable’s significance to the model, gauging the model’s overall goodness-of-fit (to determine whether a more complex model is called for), and assessing the model’s predictive power—all of which it can also use to justify the model to regulators.

Although statistically based models have historically been less common in the AML community because they appear more complex and less understandable to the layperson, they are quickly becoming the industry standard in the face of the increasing regulatory pressure to use more scientific approaches. The ability to identify the variables that contribute most to the model, the selection of coefficients (that is, weights) based on maximum likelihood estimation, the ability to assess the strength of the model, and the ability to estimate the confidence of model predictions—all these advantages favor this modeling approach. The fact that regulators prefer these models only adds to the reasons to use them.

## ORDINAL LOGISTIC REGRESSION

Once the firm has decided to move forward with a statistically based model, what type of statistical model should it select for its customer risk rating model? There are several methods to choose from, all with slightly different objectives and various strengths and weaknesses. But when the primary objective is to group customers into distinct buckets based on risk, ordinal logistic regression is a highly effective statistical modeling technique to consider. This is very often the situation a firm faces in developing a customer risk rating model, because the goal is to separate its customers into different risk groups based on their inherent risk characteristics.

Ordinal logistic regression differs from binary logistic regression in that the target variable can have more than two values and the corresponding categories are assumed to be ordered. Some logistic models allow for multiple categories of the target variable that aren't ordered (called multinomial), but ordered categories are preferred for two primary reasons:

- Ordinal models are simpler than multinomial models and therefore easier to interpret (Allison 2012).
- The hypothesis tests for ordinal models are more powerful than those for multinomial models (Allison 2012).

The cumulative logit model is the ordinal logistic regression model most commonly used; it is the default model used by SAS/STAT® software. The cumulative logit model assumes that the model can be combined into multiple binary splits of the dichotomous target variable, which is an assumption that must be initially tested using the score test for the proportional odds assumption. In cases where this assumption has been severely violated and cannot be reasonably believed to hold, usually a multinomial or binary model is used instead.

Although the cumulative logit model produces only one set of beta coefficients, the equation contains one less intercept constant than the number of target variables. The probability that an event will exist within each category is then calculated, and the category with the greatest probability is the one that the model assumes the customer belongs to (that is, this is the model’s estimated category).

The following sections walk you through the preliminary analysis that a firm must perform when it uses ordinal logistic regression to develop a customer risk rating model. By understanding the process at a high level, the firm can dispel the mystery and perceived complexity that surround these models. As a note, this paper assumes the use of SAS/STAT software; however, ordinal logistic regression is also available in other SAS® products, such as SAS® Enterprise Miner™.

### DEVELOPING AN ACCURATE TARGET VARIABLE

Before you explore the data to be used in the model, you must evaluate the target variable for accuracy. The target variable for customer risk should reflect actual historical customer experience. If the firm is not confident in the accuracy of the risk being assigned to its customers by its current model, then it must sample and review customers across the different risk levels and attribute values. These samples must be large enough to be statistically significant for the building and testing of the ordinal logistic regression model.

## MULTICOLLINEARITY

In general, multicollinearity occurs when two or more predictor variables (also known as independent variables or covariates) in a regression model are highly correlated with each other. To be more specific, multicollinearity exists when one or more of the variables used in the model can be linearly predicted within a reasonable degree of accuracy by using the other variables in the model. Note that this refers only to the relationship between the various predictive variables within the model; the predictive variables are expected to be correlated with the dependent, or target, variable.

When multicollinearity is present, the model's estimated coefficients can change erratically in response to small changes in the data or model. Although multicollinearity does not reduce the predictive power or reliability of the model as a whole, at least within the sample data used to train the model, it does affect calculations regarding individual predictions. That is, a regression model with correlated predictor variables can indicate how well all those variables predict the target variable, but it might not give valid results about any one predictor variable or about which variables are redundant.

There are two primary approaches to detecting multicollinearity within a model, and usually both are used. In the first approach, you can use the CORRELATION procedure in SAT/STAT to produce a correlation matrix to show the relationship between the predictive variables; it is important to use the Spearman correlation. The Spearman correlation coefficient is often described as being "nonparametric" and is used when data are grouped rather than numeric. The following statements illustrate the use of the CORRELATION procedure and the Spearman correlation:

```
proc corr data=YourData outs=CorrData(where=(UPCASE(_TYPE_)='CORR'))
     nomiss spearman;
  var YourVariables;
run;
```

In the second approach, you run a regression that includes all the predictive variables and request that the variance inflation factors (VIF) be provided. This produces a VIF value for each variable. The VIF is the reciprocal of one minus the coefficient of determination between the respective variable and the remaining predictor variables. Usually, a VIF greater than or equal to 4 indicates moderate multicollinearity, and a VIF greater than or equal to 10 signifies high multicollinearity. The following statements illustrate this approach by calling the REG procedure in SAS/STAT:

```
proc reg data=YourData;
  model yourvariables / TOL VIF;
run;
```

## ZERO-COUNT CELLS

Zero-count cells, or events for which there are no observations, can cause unstable results within logistic regression. Furthermore, there is no maximum likelihood estimate for the respective variable. If this problem is not addressed, the SAS/STAT procedure issues warning messages that refer to either quasi-complete or complete separation of zero-count cells. However, the messages do not specify which variable the separation exists for, so it is important to investigate this further before you fit the logistic regression model.

In exploring the predictor variables, it is important to generate cross-tabulation tables of the individual predictor variables versus the target variable in order to identify zero-count cells. In the examples throughout this paper, the firm designates customer risk by using five categories: Low, Medium, High-Low, High-Medium, and High-High. This is to facilitate stratification across its high-risk customers.

Table 1 shows an example of quasi-complete separation, where there are no Low, Medium, or High-Low risk customers that also have a Country Risk of 3.

| | | | Target Value (Risk) | | | | |
|---|---|---|---|---|---|---|---|
| Variable | Data Type | Category | Low | Medium | High-Low | High-Medium | High-High |
| Country Risk | Ordered Category | 1 | 102,528 | 117 | 268 | 68 | 31 |
| " " | | 2 | 19,337 | 181 | 57 | 28 | 28 |
| " " | | 3 | 0 | 0 | 0 | 33 | 73 |

**Table 1. Example of Quasi-complete Separation**

Table 2 shows an example of complete separation, where any customer with a Country Risk of 3 falls into the High-High customer risk category.

| | | | Target Value (Risk) | | | | |
|---|---|---|---|---|---|---|---|
| Variable | Data Type | Category | Low | Medium | High-Low | High-Medium | High-High |
| Country Risk | Ordered Category | 1 | 102,528 | 117 | 268 | 68 | 31 |
| " " | | 2 | 19,337 | 181 | 57 | 28 | 28 |
| " " | | 3 | 0 | 0 | 0 | 0 | 106 |

**Table 2. Example of Complete Separation**

You can take the following actions to handle situations of quasi-complete and complete separation:

- Remove the variable or variables causing the problem (this is an option only if the variables contribute marginally to the model).
- Combine categories if the variable contains multiple categories.
- Define a rule outside the model that automatically sets customers to high-risk when they meet certain criteria that always result in those customers being considered high-risk.
- Check to see whether another variable is a dichotomous version of the variable in question.
- Grab more sample data that reflect what is missing, if possible.

## PROPORTIONAL ODDS ASSUMPTION

The proportional odds assumption tests whether the coefficients of the dichotomous groupings of the outcome variable are the same. In ordinal logistic regression, the proportional odds assumption often does not hold. This fact is widely understood but often ignored because the practical implications, depending on the modeling objective, can be minimal. In SAS/STAT procedures, the score test for the proportional odds assumption tests the hypothesis that the estimated coefficients are not materially different from each other, regardless of the dichotomization.

Table 3 shows the mapped value of the target for the logistic regression model, and Table 4 shows the dichotomous groups that are used in the series of binary logistic regressions (that is, 1 versus 2, 3, 4, and 5).

| Target | Model Value |
|---|---|
| Low | 1 |
| Medium | 2 |
| High-Low | 3 |
| High-Medium | 4 |
| High-High | 5 |

**Table 3. Example of Target Mapping**

| Dichotomous Groups | |
|---|---|
| 0 | 1 |
| 1 | 2, 3, 4, 5 |
| 1, 2 | 3, 4, 5 |
| 1, 2, 3 | 4, 5 |
| 1, 2, 3, 4 | 5 |

**Table 4. Example of Dichotomous Groups**

The null hypothesis is that there is no statistical difference in the estimated coefficients between models. If the *p*-value is high, then the null hypothesis is not rejected, and you can conclude that the estimates are not significantly different. Output 1 shows an example of the score test results in SAS/STAT procedures. It is important to note that the score test often rejects the null hypothesis more frequently than it should. Stokes, Davis, and Koch (2012) state that this test needs at least five observations for each outcome of the category versus the target. In creating a cross-tabulation of a categorical variable versus the target variable, you need to have at least five observations in each cell. Otherwise, the sample size could be too small, there are simply no data (zero-cell problem), or it is a rare event.

| Score Test for the Proportional Odds Assumption | | |
|---|---|---|
| Chi-Square | DF | Pr > ChiSq |
| 53.4698 | 24 | 0.0005 |

**Output 1. Example of Score Test for Proportional Odds Assumption**

## MODEL DEVELOPMENT AND TESTING

In ordinal logistic regression, you must use a holdout sample to test the model to ensure that it truly fits the data and doesn't simply do so by chance, because many combinations of covariates are considered. A common practice is to build the model on approximately 70% of the data and test the model on the remaining data (the holdout data set). After testing, you can run the model on the whole data set. Then you can make comparisons between the outcome estimate percentages of the whole data set and those obtained during both model development and initial testing.

You can use the SURVEYSELECT procedure in SAS/STAT to create the build and test data sets. It contains a feature to randomly split the data. Or you can use the SAMPRATE= option to select a percentage at which to split the data.

The following statements show the SAMPRATE= option set to 0.7 (70%). The OUTALL option keeps all the records from the original data set; it creates a new variable called SELECTED that has a value of 1 if it was part of the 70% of the data and 0 if it was part of the remaining 30%.

```
proc surveyselect data=YourData out=SplitData samprate=0.7 outall;
run;
```

### VARIABLE SELECTION METHOD

SAS/STAT procedures offer several different variable selection methods—forward selection, backward elimination, and stepwise selection—to help you determine which variables to include in the ordinal logistic regression model. However, your firm might choose to forgo these selection methods and instead use specific variables that it deems important with respect to the target variable—in this case customer risk.

In the *forward selection* method, the procedure systematically evaluates each available attribute and includes the effect in the model that adds the most to model performance. Then it goes through the remaining attributes one at a time to determine whether any others significantly improve performance. The procedure terminates when no further effects can be added to the model to significantly improve performance or when all the attributes are included.

In the *backward selection* method, the procedure begins with all the attributes. The variable that has the smallest partial *F* statistic is noted. The procedure systematically evaluates each available attribute and removes the effect in the model that is the most insignificant to model performance. Then it goes through the remaining attributes one at a time to determine whether any others are insignificant. The procedure terminates when the variable that has the smallest partial *F* statistic is significant.

In the *stepwise selection* method, the procedure systematically evaluates each available attribute and includes or removes the effect in the model that adds the most to performance. Then it goes through the attributes one at a time to determine whether their removal or addition significantly improves performance. The procedure terminates when no further effects can be added to or removed from the model to significantly improve performance or when all the attributes are included.

Although these selection methods work well in a mathematical sense, firms that are developing customer risk rating models should put additional thought into selecting variables that will also satisfy regulatory expectations. For instance, if variables for high-risk geography are not included in the model, then the firm should be prepared to explain why these variables were not significant (for example, all customers are located in high-risk areas). If it cannot find a good reason to exclude the variable, it should manually add the variable back into the model.

The following statements use the LOGISTIC procedure in SAS/STAT to build an ordinal logistic regression model with forward selection by using data from the build data set, and subsequently score the test data by using the built model:

```
proc logistic data=SplitData(where=(selected eq 1))
       plots(only)=(effect(polybar)
       oddsratio(range=clip)) descending outmodel=yourModel;
  class yourOrdinalVariables / param=reference;
  model risk= yourVariables / selection=forward rsq;
  output out=yourBuildResults predprobs=individual;
run;

proc logistic inmodel=yourModel;
  score data= SplitData(where=(selected eq 0)) plots)) fitstat
     out=yourTestResults;
run;
```

## MODEL OUTPUT

Ordinal logistic regression calculates the odds ratios, coefficients, and other statistics for the significant predictor variables chosen by the model. It also calculates the probability of each risk level for a customer and assigns the risk level with the highest probability to that customer. You can use the risk assignments in turn to assess the predictive power. You can assess the predictive power of the model by using standard measures of association, which a SAS/STAT procedure can calculate for you. These predictive measures are derived from the concordant and discordant pairs observed within the data.

Figure 2 shows example estimates for the logit coefficients and odds ratios of the resulting intercepts and significant variables for an example ordinal logistic regression model. The coefficient indicates the change in the log odds ratio for each one-unit increase in the explanatory variable. For example, if a customer were to increase its SAR count by one, its risk score would be expected to increase 7.4422 units on the log odds scale, assuming that the other variables remained constant. The coefficients for the categorical variables are read differently, because the categorical values of the respective variables are compared to each other via “dummy” coding. For example, when you look at the comparison of “1 vs 3” for the variable Product Risk, you can expect a –8.693-unit decrease in customer risk as the customer moves from a product risk of 3 to 1.

The odds ratio is calculated by taking the exponent of the coefficient estimate for the respective variable. For example, the odds ratio estimate of 0.077972428 for the HIFCA Flag variable can be calculated by taking the exponent of –2.5514. This in turn can be read as the proportional odds of comparing a customer in a HIFCA area to a customer in a non-HIFCA area. In going from a non-HIFCA area to a HIFCA area, the odds of being in the highest risk level (5) are 0.078 times higher than the odds of being in the other combined lower risk levels (4, 3, 2, and 1), if all the other explanatory variables are held constant.

| Variable Name | Comparison | Odds Ratio Est | Coefficient | Standard Error | Wald Chi-square | Pr > ChiSq |
|---|---|---|---|---|---|---|
| Intercept | 5 | | 1.4041 | 2.3805 | 0.3479 | 0.5553043 |
| Intercept | 4 | | 5.1922 | 2.4435 | 4.5151 | 0.0335969 |
| Intercept | 3 | | 9.8446 | 2.5708 | 14.6642 | 0.0001285 |
| Intercept | 2 | | 14.9162 | 2.7964 | 28.4521 | <.0001 |
| Industry Code | 1 vs 4 | <0.001 | -7.9945 | 2.4146 | 10.9621 | 0.0009299 |
| Industry Code | 2 vs 4 | <0.001 | -7.3831 | 2.4181 | 9.3226 | 0.0022634 |
| Industry Code | 3 vs 4 | 0.024192805 | -3.7217 | 2.3299 | 2.5517 | 0.1101756 |
| HIFCA Flag | 0 vs 1 | 0.077972428 | -2.5514 | 0.5566 | 21.0124 | <.0001 |
| CTR Count | | >999.999 | 8.4971 | 1.7629 | 23.2307 | <.0001 |
| SAR Count | | >999.999 | 7.4422 | 2.2739 | 10.7118 | 0.0010645 |
| Product Risk | 1 vs 3 | <0.001 | -8.693 | 1.1207 | 60.1694 | <.0001 |
| Product Risk | 2 vs 3 | 0.027171137 | -3.6056 | 0.5698 | 40.0372 | <.0001 |
| Expected Cash | | 4.176192724 | 1.4294 | 0.2097 | 46.4574 | <.0001 |

**Figure 2. Example Summary Table**

Output 2 shows example estimates of the predictive power of the model (the values range from 0 to 1, with larger values signifying greater predictive power). The pseudo-coefficient of determination, often signified as R-square ($R^2$), is another popular statistic that you can use to assess the predictive power of the logistic model, where the Max Rescaled R-Square adjusts the statistic to account for the fact that in a discrete outcomes model the R-square value can often never actually equal 1.

The LOGISTIC Procedure

Probabilities modeled are cumulated over the lower Ordered Values.

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 96.9 | Somers' D | 0.947 |
| Percent Discordant | 2.2 | Gamma | 0.955 |
| Percent Tied | 0.9 | Tau-a | 0.708 |
| Pairs | 5705 | c | 0.973 |

| Fit Statistics for SCORE Data | | | |
|---|---|---|---|
| R-Square | Max-Rescaled R-Square | AUC | Brier Score |
| 0.758697 | 0.800536 | . | 0.308647 |

**Output 2. Example Ordinal Logistic Regression Results**

In general, tests of the model's predictive power assess how well you can predict the target variable by using the covariates (that is, predictive variables). It is entirely possible to have a model that predicts the target variable very well but fails the goodness-of-fit tests. It is also possible to have a model that makes poor predictions but shows very good model fit. Predictive power is commonly calculated using the measures of association of Somers' D, gamma, tau-*a*, and c (or AUC), which are all listed in Output 2.

Another useful way to view the model results is to generate a two-way contingency (cross-tabulation) table of the predicted target versus the actual target, as shown in Figure 3.

| Model Error Severity (Combined Data) | | | | | | |
|---|---|---|---|---|---|---|
| Estimated Target | Actual Target | | | | | Row Total |
| | Low | Medium | High-Low | High-Medium | High-High | |
| Low | 57 | 1 | 1 | | | 59 |
| Medium | 2 | 47 | 4 | | | 53 |
| High-Low | | 1 | 32 | 13 | 2 | 48 |
| High-Medium | | | 2 | 18 | 3 | 23 |
| High-High | | | | 3 | 25 | 28 |
| Total | 59 | 49 | 39 | 34 | 30 | 211 |

| Field Key |
|---|
| Correct Prediction |
| Inaccurate by 1 |
| Inaccurate by 2 |
| Inaccurate by 3 |
| Inaccurate by 4 |

| Error Rates | Count | Percent |
|---|---|---|
| Correct Prediction | 179 | 84.83% |
| Inaccurate by 1 | 29 | 13.74% |
| Inaccurate by 2 | 3 | 1.42% |
| Inaccurate by 3 | 0 | 0.00% |
| Inaccurate by 4 | 0 | 0.00% |
| **Total** | **211** | **100.00%** |

**Figure 3. Example Contingency Table**

You can also use the LOGISTIC procedure to produce an output data set that contains the predicted customer risk along with the predicted probabilities for each level of risk. This is very useful because firms usually want a score that is associated with the risk level assigned to each customer. A score can be calculated by using the sum of the weighted probabilities.

If you have five risk levels (1–5), the score would fall between 1 and 5. However, you can apply a scale by multiplying the weighted probability of the score by some factor. In Figure 4, Customer X would be assigned a risk level of 5 (High-High) because it has the highest calculated probability. To calculate the weighted probability, you add the weighted probabilities together. Because you have five risk groups, multiplying the sum of the weighted probabilities by 20 results in scores from 20 to 100. Figure 4 shows an example of these numbers.

| | Customer X | | | | |
|---|---|---|---|---|---|
| | Low | Medium | High-Low | High-Medium | High-High |
| Probability | 0.000000000 | 0.000000038 | 0.000004454 | 0.000943457 | 0.999052050 |
| Weight | 1 | 2 | 3 | 4 | 5 |
| Weight * Probability | 0.000000000 | 0.000000077 | 0.000013361 | 0.003773830 | 4.995260252 |
| Weighted Probability | 4.99904752 | | | | |
| Scale to 100 Points | 99.98095039 | | | | |

**Figure 4. Example of Scoring**

### DEPLOYMENT

During the deployment step, the model logic is implemented in the operational production system. There are many ways to do this, including SAS web services, batch SAS processing, various queuing servers, and recoding the model logic into the language that the operational production system expects. For this process, it is assumed that the model will be deployed as a SAS batch process and that input data delivered by the firm match the data delivered for the modeling process.

## MODEL RISK GOVERNANCE

When the customer risk rating model is created and put into operation, the firm must create a plan for ongoing model validation, as described in *Supervisory Guidance on Model Risk Management* (OCC 2011-12). Deliverables might include the validation of the target variable used to train the model, the validation of the model performance, and a model validation report.

Firms might want to perform analysis to assess the relationship between the target variable (Customer Risk Rating) and the actual resulting number of scenario alerts generated, case referrals, or Suspicious Activity Reports (SARs) filed on the customer. This enables the firm to document that customers receiving a Customer Risk Rating of 5 (High-High) are more likely to be involved in suspicious activity than customers that receive a rating of 4, 3, 2, or 1. If it is determined that the target variable lacks the desired degree of accuracy, then the firm has the option to update the target variable setting and retrain the model on the revised data set. This would provide a model that more accurately identifies customers that meet the firm's definition of "risky."

Regulators expect all analytical models to be validated on an ongoing basis, as described in OCC 2011-12. This process involves tasks such as periodically assessing of the model's performance, ensuring that appropriate model controls are in place, determining that the right covariates are included in the model, adjusting the coefficients as needed, and so on. It can also include testing new variables that were not available before or that contained only sparse data.

Firms should also produce a model validation report every year that documents *all* model validation tests that it has performed and the results of those tests. The report should also contain the following:

- a description of the model, its parameters, its input variables, and its strength and weaknesses
- validation of all model components, including input data, assumptions, processing, and reports
- an evaluation of the model's ongoing conceptual soundness, including development details that might be relevant
- evidence of ongoing monitoring, including process verification and benchmarking
- an outcomes analysis, including back-testing

## CONCLUSION

In addressing customer risk rating by financial firms, SAS has found that statistically based models are the most effective method of classifying customer risk while satisfying the regulatory expectation that quantitative modeling techniques and practices be used. Although ordinal logistic regression modeling requires a variety of analyses and testing, this approach has been accepted in the Anti-Money Laundering community as a successful statistically based method of measuring customer risk. The key to any AML modeling is to ensure that development, testing, and validation are well documented and explainable to regulators. Increased regulatory pressure requires continuous evaluation of a firm's customer risk. To meet these demands, the AML community has turned to analytical and statistical methodologies to improve customer risk assessment in order to effectively manage risk and identify the customers that require the most attention and review.

## REFERENCES

Allison, P. D. (2012). *Logistic Regression Using SAS: Theory and Application*. 2nd ed. Cary, NC: SAS Institute Inc.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2012). *Categorical Data Analysis Using SAS*. 3rd ed. Cary, NC: SAS Institute Inc.

## CONTACT INFORMATION

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

Edwin Rivera
SAS Institute Inc.
Edwin.Rivera@sas.com

Jim West
SAS Institute Inc.
Jim.West@sas.com

Carl Suplee
SAS Institute Inc.
Carl.Suplee@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.

# Ready to take your SAS® and JMP® skills up a notch?

Be among the first to know about new books, special events, and exclusive discounts.
**support.sas.com/newbooks**

Share your expertise. Write a book with SAS.
**support.sas.com/publish**

sas.com/books
*for additional books and resources.*

sas
THE POWER TO KNOW®

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. © 2017 SAS Institute Inc. All rights reserved. M1588358 US.0217