

Creating Clarity Out of Chaos: An Exercise in Defining Actionable Intelligence

Christopher Westphal
Visual Analytics Inc.
20010 Fisher Avenue, #200
Poolesville, MD 20837
www.visualanalytics.com

Abstract

Chaos is a common term used to represent extreme confusion, disorder, and disarray. Much of the intelligence production process used throughout law enforcement, the intelligence community, and other investigative organizations is based on accessing, combining, analyzing, and reporting on data that comes from multiple sources in different formats spanning across many systems. The chaos introduced from variations in content, formats, spellings, errors, and other incompatibilities makes it very difficult to understand and detect the patterns in the data. As new systems and technologies are introduced into the marketplace, they offer better methods to standardize, classify, clean-up, disambiguate, and provide better clarity on the data. This paper discusses a number of chaotic factors associated with data and novel techniques and approaches for clarifying their use in proactive analytics with a specific focus on the use of network diagramming and visual charting.

Introduction

Government agencies, businesses, and industry have dedicated enormous resources to the collection and storage of information. However, only a small amount of this data is actually used due to complexity and volume. The transformation of relevant information into a form that enhances decision-making can be difficult, if not impossible in many cases. Advanced data mining and business intelligence technologies hold great promise for managing large, multi-format application environments because they can be designed using standard interfaces and communication formats. The application of these technologies allows the users to examine larger quantities of information with greater accuracy and speed by taking data from multiple sources and presenting it in formats that can be easily understood. Data mining methods can be used to help users discover new correlations or

associations quickly to reveal patterns and trends previously hidden in the data. This allows users to increase the probability of a successful analysis. More importantly, effective interpretation of data significantly enhances productivity and provides more timely and accurate information.

In the wake of all the corporate fraud, terrorist events, money laundering, and other wastes/abuses in our society, much of the discussions in this paper are based on the detection of fraud, criminal activity, and other "bad person" behaviors. Improving the quality of data and analysis also enhances pharmaceutical research, sales and marketing, public safety, science, and many other domains that are not criminally oriented.

Many times, organizations simply do not have the means or resources necessary to check all the available data sources to look for inconsistencies or indicators of fraud and abuse. This is not only true for commercial businesses but also for government agencies, especially law enforcement. The following sections provide some representative types of data that can be easily integrated into standard review protocols or criteria. Although these sources of data actually exist, they still must be acquired, configured, and incorporated for use within an appropriate analytical environment. Proper database structure and indexing will impact the utility of each source discussed.

Prisons or Detention Facilities. Imagine a bank receiving a loan or credit card application in the mail from a convict or inmate at a state or federal prison. If the circumstances are known, the application will most likely be denied. However, if the applicant does not reveal that they are incarcerated, the bank may never be the wiser. Addresses can appear as legitimate businesses, military bases, or other facilities and

the name of the applicant and his/her inmate number can be disguised as a mail-drop or room number (e.g., John Doe #12345). Furthermore, the name of the prison could be removed if it was being portrayed as a home address.

One simple check would be to compare the address listed on the application to a list of known prisons. For example the following represent actual prison addresses:

ASP - Fort Grant
P.O. Box 2500
Fort Grant, AZ 85644

ASP - Fort Grant stands for the Arizona State Prison Complex Fort Grant which is a vocational prison for male inmates.

BWCI
660 Baylor Boulevard
New Castle, DE 19720

BWCI stands for the Baylor Women's Correctional Institution

CDC
P.O. Box 1011
Imperial, CA 92251

CDC stands for the California Detention Center (CDC)

Camp Hawthorn
P.O. Box 140
Kaiser, MO 65047

Camp Hawthorn is a satellite facility associated with the Ozark Correctional Center for male inmates.

As can be seen in these examples, cross-checking addresses can expose situations that are suspicious, questionable, or undesirable depending on the nature of the address listing. Additionally, consider addresses from public-type locations including:

- Bus depots/terminals
- Landfills/Refuse sites
- Recycling centers
- Taxi cab companies
- Bars/Pubs
- Impound lots
- Laundromats
- Gas stations

Social Security Death Master (SSDM). The Department of Commerce offers this database with over 95 million records. Simply, it contains the name, social security number, date of birth and death, and some basic address data on any person who has died with an assigned social security number.

The following sample shows the basic layout for a record in the death master file:

- Code A/C/D
- SSN
- Last Name
- Name Suffix
- First Name
- Middle Name
- Code V/P
- Date of Death
- Date of Birth
- State/Country of Residence
- ZIP Code of Last Residence
- ZIP Code of Payment Benefit

Once a person is reported deceased, their SSN should no longer be used for any type of income reporting, finance application, or other type of official benefit. Although not authorized, use of the SSNs of the deceased for these types of purposes occurs on a fairly routine basis.

Many of the financial intelligence units operating in the United States utilize the SSDM to cross-check cash transaction reports and other types of financial dealings to ensure people are not listing the SSNs of dead people. Often, through data entry, character transpositions, or illegible handwriting (e.g., 4 looks like 9, 2 looks like 5, 1 looks like 7, etc.), these errors can occur. Additionally, intentional misrepresentations occur when people make up their own SSNs to avoid government reporting and filing requirements.

Telephone Subscriber Data. Often the data collected by organizations and governments include some type of input for telephone numbers. Whether you are making travel reservations, transferring money, or applying for credit, there is almost always a place to enter a phone number. Many people take the number provided for granted because it is seldom checked or used and usually only if there is a problem (e.g., a flight was canceled; a bank's routing numbers were wrong). Generally, people will provide some "valid" means to be contacted under these circumstances.

However, we all know that phone numbers can change quickly, get reassigned to other people, or can be unlisted or blocked. This can often raise questions with respect to their reliability in the context of performing analysis. Luckily, simple checks can be performed to help identify inconsistencies or questionable data.

In one particular state, the Attorney General's Office subpoenas all of the public pay phones installed and operating throughout the state on a yearly basis. This produces a fairly basic data set where the phone number, operating company, and physical location of the phone are listed. The data provides little value by itself since all are operated by legitimate companies. However, it becomes invaluable when using this data in conjunction with wire-transfers, job applications, or other situations where "truth" and cooperation from a person are considered important. Imagine a police officer filling out an arrest report where the phone number provided by the suspect is for a public pay phone at the corner gas store. A simple check would expose this fact and appropriate measures could then be taken to acquire legitimate data.

Other checks using phone numbers can be made using more public information such as phone books. Sources such as Google have provided the convenience of incorporating the white pages into their search engine databases. For example, if you search on a 10-digit number, formatted or unformatted, Google will prioritize the matches to deliver the phone number, subscriber, and address as the top-returned item on the result list.

The name of the subscriber can be compared against the given name and if a mismatch occurs, follow-up inquiries can be initiated. In many of the analytical systems deployed within the financial and law enforcement communities, this check can prove vital for detecting inconsistencies or incomplete data. In fact, it often identifies additional targets that were not previously known.

Furthermore, once an address is provided, there are secondary checks that can then be made with respect to the original data. Thus, the phone number leads us to a subscriber and an address – and if these do not match the original data, it warrants additional investigative time to clarify the inconsistencies. It also gives investigators additional dimensions to verify, especially the distance between known

addresses and other subjects with the same address or phone number.

FBI's Most Wanted. Besides the obvious reference to Usama Bin Laden on the FBI's Top 10 Most Wanted site, how many people know who the other 9 individuals are? What businesses routinely check the monthly fugitive list to determine if they are dealing with a criminal? How many of us know even if they are living next door as our neighbors? Often these types of checks or inquiries are never performed because the convenience of accessing the information is not readily available.

Many credit checks, money transfers, job applications, school transcripts, and other references are unknowingly conducted for wanted felons. Beyond what is listed on the FBI's Web site, there are many others, including:

- Marshals Service
- Secret Service
- Drug Enforcement Administration
- Alcohol, Tobacco, and Firearms
- Royal Canadian Mounted Police
- Interpol (International)

Furthermore, state and local law enforcement have extended watch and be-on-the-look-out (BOLO) lists for a number of critical offenses. Many states are also posting their sex-offender registries online, offering another source of reference for checking on someone.

There are large numbers of sources available in the public domain for searching. Subscription services like LexisNexis and ChoicePoint provide a consolidated view from a number of sources including credit headers, asset location, UCC filings, real property, bankruptcies, and telephone directories. The benefit of subscription services is that they have done the work of accessing and consolidating the data (nationwide) into a single service. Additionally, they usually run some clean-up routines to provide consistent names, addresses, and an overall profile of the search targets.

Other sources of public records, compiled by companies such as Search Systems (public records), provide a large number of "accessible" sites broken down by state and nationwide (U.S.) access that can be searched through online interfaces.

For example, in the state of Maryland, there are a number of public sources that can be accessed over the Internet. Some examples of these sources include:

- Barber Licenses
- Certified Public Accountants
- Charitable Organizations
- Cosmetologist Licenses
- Inmates (State)
- Nursing Homes
- Pawn Brokers
- Plumbers
- Real Property Records
- Sex Offenders

The only draw-backs to this type of approach is that you need to know what is available and you have to manually visit each site and use a different interface to generate a query for each target of interest. Although extremely useful, it does not lend itself to operational systems with large scale analytical demands.

Meta-Data

There are a lot of data in data. Meta-data is seldom used in analytical systems because of certain sensitivities (algorithms are not public knowledge) or because they are unknown to the analyst. Providing better utilization of the meta-data can make a big impact on the end results and the detection of critical patterns. The following describe some of the more common forms of meta-data utilized within the intelligence community.

Dates – It may seem a little unusual to think of a “date” as containing meta-data. However, the use of dates can provide incredible insight into investigations pertaining to money laundering, organized crime, and terrorism. Generally, dates are only interpreted at “face-value” such that 03/11/2004 does not contain much meaning except that it is March 11, 2004.

The meta-data for this date, although not explicit in its basic representation as DD/MM/YYYY, contains the following:

- Year = 2004
- Month = March
- Day of Month = 11
- Day of Week = Thursday
- Week of Year = 11
- Day of Year = 71
- Season = Winter
- Holiday/US = No

From here, we can cascade meta-data about other events as well. Since 03/11/2004 is not a holiday in the United States, and therefore represents a standard work day, we can reason that banks will be open. We also know that if banks are open, individuals and businesses will actively transact money at these institutions – loans will be approved; accounts set up and closed; and a myriad of other related items will occur. When possible, meta-data should be used help perform analysis.

Furthermore, the meta-data can represent the occurrence or anniversary of significant events. Some represent “absolute” dates, just as January 1st always correlates to New Years Day and February 14th is Valentines Day. Other dates are “relative” such that they do not always happen on the same exact date; Mardi Gras (Fat Tuesday) and Thanksgiving Day (third Thursday in November) represent examples of relative date events.

The intelligence community is always vigilant about world events and different dates are more or less important depending on the type of analyses and where in the world it is being conducted. The date, 03/11/2004 correlates to the train bombings in Madrid, Spain. Incorporating these types of references and lookups within an analytical system can help correlate and expose patterns of interest.

Additionally, we can perform calculations relative to other dates. For example, our date of discussion represents 911 days since September 11, 2001. Knowing this information can help provide the intelligence community additional insights into predicting or exposing similar events.

Check Digits – Many types of numbers are encoded with special values used for description or validation purposes. Often there is a check-digit or a combination of values that can be used to certify if the number is structurally accurate. Sometimes the validation is based on a mathematical formula or simply through a lookup table. The following examples provide a high-level overview of several identification numbers commonly used in everyday transactions and activities.

SSN Validation – Many data sources contain Social Security Numbers (SSNs) as a form of identity or descriptive information about an individual. However, there are instances where

SSNs are not properly entered or they are made up by people to avoid being tracked. Luckily, the process used to create SSNs helps isolate certain types of abuses from occurring.

As many people know, an SSN consists of nine (9) digits. The first three (3) digits indicate the state or region where the SSN was issued. The following examples show how SSNs are related to their issuance locale:

050-134 New York
261-267 Florida
526-527 Arizona
545-573 California
530 Nevada
574 Alaska
580 Virgin Islands

The next two digits (the middle 2 numbers) are called the “group” and reflect a specific ordering used to help validate whether the SSN was ever issued. The group number is assigned using a sequence of odd numbers for 01-09 and even numbers for 10-98. Initially the numbers are assigned as follows:

odd numbers: 01 to 09
even numbers: 10 to 98
even numbers: 02 to 08
odd numbers: 11 to 99

The group numbers are validated using a “high” number that is posted on the Social Security Administration Web site and updated monthly. This high-number list indicates the more recent group assignments for each state/region code. Additionally, group codes of “00” are not assigned and considered invalid.

Finally, the last four (4) digits are consecutively assigned 0001-9999 and do not have any sort of check digit or lookup values. Keep in mind, this algorithm is only used to determine if an SSN has ever been issued – it can’t tell if the person is alive or dead or the name of the person to whom the number was originally issued.

VIN Validation – Since 1981, all vehicle identification numbers (VINs) have been standardized to 17 characters that encode specific details about the vehicle. Knowing this information can help verify the legitimacy of the vehicle, especially for insurance-related applications or law enforcement check-points at border crossings. Each character represents a

unique aspect about the vehicle as shown below:

1	Country
2	Manufacturer
3	Make
4-6	Engine
7	Body/Transmission
8	Trim Level /Restraint
9	Check Digit
10	Model Year
11	Assembly Plant
12-17	Serial Number

The Check Digit (position #9) is based on a mathematical calculation. Generally, each character in the VIN is assigned a number, which in turn is multiplied by a position-weight-factor as defined in a standardized lookup table. The products are then added together and the total divided by 11. The remainder becomes the check digit (the value 10 = X). Without the proper knowledge of each VIN value or how the check digit is calculated, it is difficult to just make up a fake VIN.

Luhn Validation – Credit card numbers are often encoded with certain information used to identify the credit card company, the issuing bank (financial institution), the account number, and other information.

For example, American Express cards are 15-digits long and start with either 34 or 37. MasterCard numbers are 16-digits and start with a number between 51 and 55. Visa numbers can be either 13- or 16-digits long and always start with the number 4. Diners Club, Discover, and many other popular cards also support this type of encoding with their numbers.

Credit card numbers also have a built-in check digit to verify that the number is valid. Most credit card companies employ the Luhn formula (also known as Mod-10) for calculating the check digit. This technique was created back in the 1960s and is also used to validate other non-credit card numbers such as Canadian Social Insurance Numbers and other financial services numbers.

The following example shows how to apply the Luhn formula using the sample credit card number:

541234567890125

Note: The last digit shown and underlined (5) is the check digit.

First, double the value of alternate digits (shown underlined) beginning with the second to last digit from the right as shown below:

$$\begin{array}{r} 5 \underline{4} 1 \underline{2} 3 \underline{4} 5 \underline{6} 7 \underline{8} 9 \underline{0} 1 \underline{2} 5 \\ \times 2 \quad \times 2 \quad \times 2 \quad \times 2 \quad \times 2 \quad \times 2 \quad \times 2 \\ \hline 8 \quad 4 \quad 8 \quad 12 \quad 16 \quad 0 \quad 4 \end{array}$$

Add the individual digits comprising the products to each of the unaffected digits in the original number. If the number created represents a double digit, then the product is each of the parts. For example 18 will become 1 + 8 = 9.

The following additions are performed on the above sequence:

$$5+8+1+4+3+8+5+(1+2)+7+(1+6)+9+0+1+4+5= 70$$

The resulting product must be a number ending in zero (so there is no remainder) for the account to pass the modulus 10 test and be valid. In this case, the sample number passes the Luhn algorithm. Keep in mind that just because a number passes this basic "validation" technique does not necessarily mean it is truly valid (e.g., issued). All it means is that it conforms to the sequence and pattern followed by the issuer. Thus, taken at face value, these approaches can detect the obvious mistakes, invalid attempts, or other blatant errors to help validate the data.

Examples

The following examples discuss creating actionable intelligence out of a wide number of disparate data sources. Often, the databases accessed have different formats, representations, and contents and many are stored on different platforms with different security protocols. The goal is to create a clear and concise picture of all the events, relationships, and associations contained in the data.

Keep in mind that the technologies necessary to access, integrate, analyze and present multi-source data already exist and are in operation throughout the world. The evolving capabilities have more to do with redefining the business intelligence and rules by which to expose the desired patterns and trends. In many cases, the intelligence, law enforcement, and investigative

communities do not know what important indicators are or what combinations of conditions are critical for detecting terrorist activities or money laundering operations.

An important goal in proactive analytics is to help expose the anomaly, inconsistency, or mistruth in the information. To be fair, just because someone gets associated with an SSDM (death master) entry, whether intentional or not, does not prove they are naughty or nice. Rather, it provides a discriminator among the data to identify well-qualified targets for which additional investigative resource can be applied.

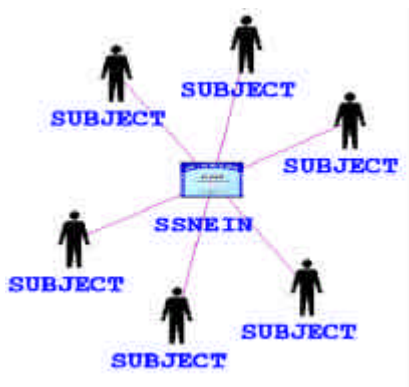
In the context of the data analysis, there also must be provisions in place where "bad" data can be corrected or overridden to avoid false positives. In many cases, simple value transpositions, encoding errors, or content variations can cause undesired consequences.

The following examples are based on actual link diagrams derived from data routinely used in exposing money laundering and financial crimes. Millions of Bank Secrecy Act (BSA) forms are filed every year ranging from Cash Transaction Reports (CTR) for amounts over \$10,000 to Suspicious Activity Reports (SAR) for questionable money movements at any dollar amount. This immense BSA data collection with over 300 million records stored in a dozen databases has virtually every data inconsistency imaginable including misspellings, null values, and data entered into the wrong fields (e.g., a last name entered into a city field). To make matters worse, the "suspects" are not always truthful when presenting information to the financial institutions.

Needless to say, the large degree of chaos contained in this data represents a major challenge to overcome. Exposing the networks, connections, and details for each of the objects is crucial to interpreting, understanding, and clarifying what the data truly means.

Example #1 – In this first example, the network structure presented is generally called a starburst and depicts a single SSN connected to six (6) SUBJECTS. Initially, most investigators would consider this a very questionable and suspicious situation, especially if the SSN appeared valid and the names of each SUSPECT were different.

An alternative interpretation of this network would be appropriate if the SSN represented an invalid or common number such as 999999999, 000000000, UNKNOWN, or NOT PROVIDED. Often in this type of financial data, there is a lot of dirty data. If this were the case, the entire network would be disregarded because there is no reliable connection among the SUSPECTS.



A different scenario would arise if each of the SUSPECTS had a similar name. In this case the investigators may discount the severity of the pattern if the names represented the following:

- John Smith
- Johnny Smith
- J. Smith
- Jon Smithe
- Juan Smith
- J J Smith

Obviously, these names all reflect the same SUBJECT. Of course the question then becomes: Is this person trying to avoid detection by varying his name? It would be highly unusual for each transaction to have a different name spelling. Also, a thicker link in between SUSPECT and SSN would indicate a repeated usage of the same name; however, none appears in this particular network. Therefore, the pattern generates a stronger interest from the investigators.

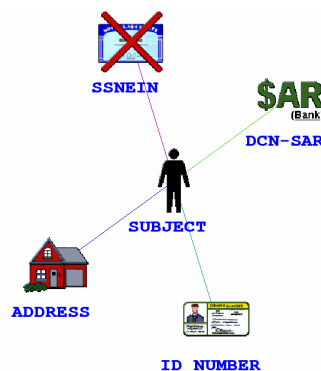
What is not explicitly conveyed in the diagram is that each linkage between a SUBJECT and an SSN is generated based the occurrence of a separate and unique financial transaction. Thus, there are at least 6 wire transfers involved in creating this network, if not more.

Example #2 – From the previous discussions regarding the use of reference data such as prison addresses, public pay phone numbers, or

criminal citations, this example presents how the SSDM (death master) data can expose questionable results.

The following diagram shows the most basic information derived from a single Suspicious Activity Report (SAR). The SUBJECT is linked to an ADDRESS, ID NUMBER, SSN, and the SAR.

In this case, a special icon is displayed when the SSN from the transaction is matched to a value



appearing in the Social Security Death Master (SSDM) database. The red X overlaying the icon clearly shows the SSN is derived from a dead person.

There are only three (3) ways this situation can happen: mistakenly, intentionally, or naturally. Many times the data entered is transposed and number such as 2 and 5, 4 and 9, and 1 and 7 can be misinterpreted when poor penmanship is involved. This represents one of the most common reasons for false-positive death master matches. This is often quickly resolved because there is only a single link to the number and many times the real SSN is also displayed when multiple transactions are recorded.

Intentional use of a dead person's SSN is observed when there is repeated use of the number, depicted as a thicker link between the SSN and the SUBJECT. It is very unlikely that the same SSN would be misrepresented multiple times if it was used in many different transactions.

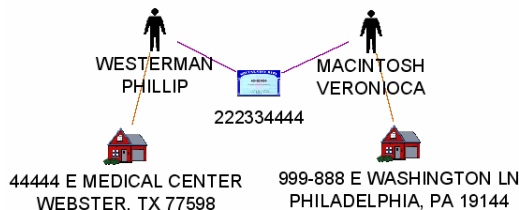
Finally, people do die. If the transaction date of the SAR occurs before the date-of-death listed on the SSN, then the SUBJECT is regrettably deceased. One of the first checks an investigator makes is to compare the respective dates to determine the sequence of events.

Example #3 – This example builds on the concept of shared values among the same entities. Again, an SSN is depicted in use by two (2) SUSPECTS. In this network, the actual names are shown for each of the objects presented.



Sometimes in the financial data, an SSN will be shared by a husband and wife in certain types of transactions. In this case, the names are not even close, so the investigators consider these people as unrelated.

At this point, the investigators want to know why both SUSPECTS are using the same SSN. The network is expanded to show the ADDRESS for each, as shown below.



Addresses are perhaps the most widely varying data encountered in any system. There are many abbreviations, spellings, and formats used to encode an address. It is not unusual to see 3, 4, or 5 variations of the same ADDRESS – often differentiated only by extra periods, commas, or directional encoding (e.g., NW, N., or North).

For the two ADDRESSES shown in the diagram, the investigators quickly see they are not even close to one another. If they were in the same CITY or STATE, there would be more of a chance the SUSPECTS were related. Unfortunately, these two addresses are more than 1,300 miles apart from one another – which dramatically diminishes the likelihood they are related.

From here, the network is expanded to show other tangible entities to include PHONES and ID NUMBERS in the hopes of finding a secondary connection among the SUSPECTS.



The premise being that a common phone number or a shared driver's license in conjunction with the SSN would guarantee a strong connection between the two SUSPECTS. Yet again, there is no additional overlap.

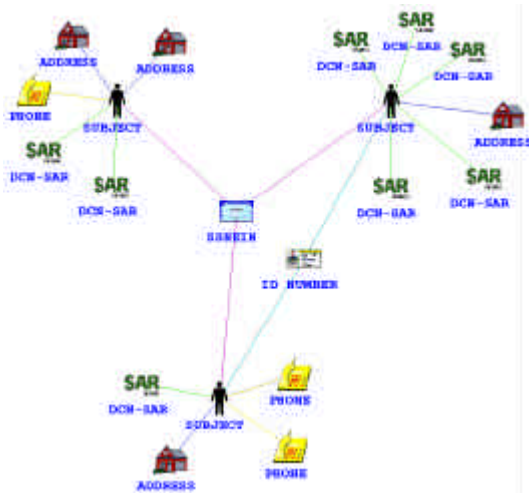
Finally, the financial transactions are displayed in the network.



As shown in this final diagram, each SUSPECT has only a single, unique transaction (SAR). This tells the investigators the SUBJECTS are not actively engaged in multiple transactions and therefore the common SSN is most likely a data entry problem and the entire network can be discounted.

If the SUBJECTS each had more than one transaction, it would be highly unlikely that the same transposition would occur for every transaction. If that were the case, the investigators would aggressively pursue these SUBJECTS.

Example #4 – This example presents a more complex network depicting the intentional use of the same SSN and ID NUMBER by different SUSPECTS. Although the diagram shows a similar structure to the previous example, the main distinguishing factor is that each SUSPECT has multiple transactions or commonality.

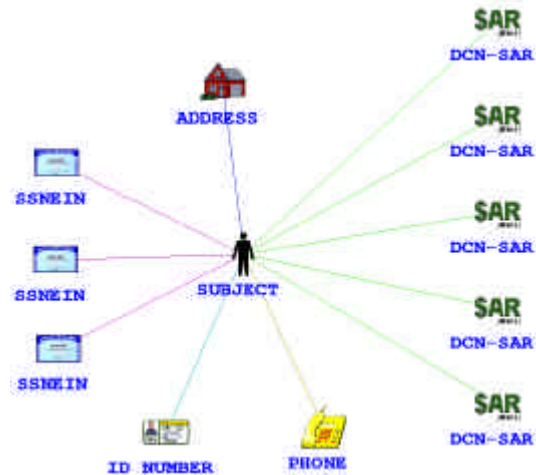


The upper-left SUSPECT was involved in two (2) transactions with a single SSN. In this case, both transactions used the same SSN and the two ADDRESSES are the same with just a slightly different spelling (AVE / AVENUE).

The upper-right SUSPECT has five (5) transactions all using the same SSN. Additionally, there is an ID NUMBER that is shared with the SUBJECT at the bottom along with the same SSN.

This network clearly shows the intentional use of the same SSN by different SUSPECTS. In fact, the ADDRESSES listed for these SUSPECTS are in different cities indicating they are operating as an organized group.

Example #5 – In this final example, a single SUSPECT is shown with connections to three (3) unique SSNS that were generated from five (5) different financial transactions (SARS).



This situation reflects the opposite of the previous examples where multiple SUSPECTS used the same SSN. Instead, it is a single SUSPECT using multiple SSNS.

As we have learned, the variations in the SSNS could be innocent data entry errors or they could be intentional misrepresentations. The most unusual factor in the diagram is that there is only one ADDRESS, PHONE, and ID NUMBER. The investigators find this inconsistent with the number of SSNS shown and therefore determine the SUSPECT is trying to avoid detection by altering his SSN.

When using object representations for network diagramming, the uniqueness of, say, a SUBJECT, is limited to the combination of first-, last-, and middle-name. Therefore, common names (e.g., John Smith, Tran Nguyen, and Mohammed Fayyad) may produce a composite representation of more than one SUBJECT. In these cases, it is common to see multiple SSNS, however, there will also be multiple ADDRESSES or PHONES or ID NUMBERS.

Conclusion

Accessing billions of records across thousands of databases provides a very challenging environment from which to conduct analysis. Making sense of all this data can be overwhelming –especially with the amount of variation in content and representation differences, to name a few of the major challenges. Often the data is not readily available for real-time, distributed, or batch access or its format does not conveniently fit the existing analytical models.

Many times there are data about data that can be exploited if known. This meta-data is usually stored in look-up tables or can be computed by simple procedures or algorithms. The value-added to the primary data source through the application of meta-data can be invaluable and

the difference between success and failure for an investigation.

The analytics applied to the masses of data help standardize and clarify their contents. Systematically reviewing how information is connected, exposing too many or too few of certain items, or simply cross-referencing values with other data is the key to understanding the data and exposing the patterns.

Realize that there are always exceptions to the patterns, and there are always exceptions to the exceptions. There is no substitute for a seasoned investigator and many of the patterns can only be exposed through their involvement, their iterations through the data, and questions they raise interpreting the results.

References

The following URLs provide references to several data sites:

FBI	http://www.fbi.gov/mostwanted/topten/fugitives/fugitives.htm
Interpol	http://www.interpol.int/Public/Wanted/Search/Recent.asp
ATF	http://www.atf.gov/wanted/
Marshals Service	http://www.usmarshals.gov/investigations/most_wanted/index.html
DEA	http://www.usdoj.gov/dea/fugitives/fuglist.htm
Secret Service	http://www.ustreas.gov/usss/advisories.shtml
RCMP	http://www.rcmp.ca/wanted/index_e.htm

Social Security Numbers info:

<http://www.ssa.gov/employer/stateweb.htm>
<http://www.ssa.gov/employer/highgroup.txt>
<http://www.ssa.gov/employer/ssnweb.htm>
<http://www.ssdmf.com/>

A large collection of publicly accessible data (by state) can be found at:

<http://www.searchsystems.net/>

Public records subscription services are offered by:

<http://www.choicepoint.com/business/public/public.html>
<http://www.lexisnexis.com/risksolutions/lawenforcement/features.asp>

Vehicle Identification Numbers (VIN)

<http://www.access.gpo.gov/nara/cfr/waisidx/49cfr565.html>
<http://www.verifyvin.com/>

Prison Addresses

<http://www.prisonerlife.com/prisondirectory.cfm>