identitysystems™

Using Identity Resolution to Enhance Your Anti-Money Laundering Program

By David Loshin, President, Knowledge Integrity, Inc.

A white paper sponsored by Identity Systems
January 2007

Contents

Anti-Money Laundering

As a result of the passage of the USA PATRIOT Act following the events of September 11, 2001, federal law mandated private organizations to take steps in identifying and preventing money laundering activities that could be used in financing terrorist activities. Conceptually, money laundering encompasses two forms of activities — performing transactions using money that is the result of criminal activity in order to hide its illegal origins, or using money for criminal purposes. More formally, the prohibition in 18 USC § 1956 specifically focuses on defining money laundering associated with financial transactions targeting:

A) 1) Whoever, knowing that the property involved in a financial transaction represents the proceeds of some form of unlawful activity, conducts or attempts to conduct such a financial transaction which in fact involves the proceeds of specified unlawful activity:

   a) i) with the intent to promote the carrying on of specified unlawful activity; or
      ii) with intent to engage in conduct constituting a violation of section 7201 or 7206 of the Internal Revenue Code of 1986; or

   b) knowing that the transaction is designed in whole or in part:
      i) to conceal or disguise the nature, the location, the source, the ownership, or the control of the proceeds of specified unlawful activity; or
      ii) to avoid a transaction reporting requirement under State or Federal law

While assessment of potential money laundering activities historically had centered on analyzing past transactions to determine if the activity had already taken place, the implication of the PATRIOT Act is that organizations must take a proactive approach to enforcement. And while one might assume that complying with anti-money laundering (AML) statutes is confined to banks, any financial institution as defined by the Bank Secrecy Act (and in 31 USC § 4312), including currency changers, insurance companies, dealers in precious metals, pawnbrokers, loan companies, travel agencies, real estate businesses and casinos, among others, is charged with monitoring customer activity for money laundering.

The protocols of AML imply a few operational perspectives: establishing policies and procedures to detect and report suspicious transactions, ensuring compliance with the Bank Secrecy Act, as well as providing for independent testing for compliance to be conducted by outside parties. But in essence, AML compliance revolves around a relatively straightforward concept: knowing your customer. Because all monitoring centers on how individuals are conducting business, any organization that wants to comply with these objectives must have processes in place for customer identification and verification. And this process does not stop once an account is established, but should extend over the lifetime of the customer relationship:

- Verify the identity of any individual who is in any way involved in establishing an account,
- Maintain clean and consistent records of the data that is used to verify a customer's identity,
- Maintain a track record of any customer's activity to flag any suspicious behavior, and
- On a continuous basis, ensure that the customer does not appear on government lists of known or suspected terrorists, or belong to known or suspected terrorist organizations.

Really Knowing Your Customers

"Identity crimes, such as money laundering, cost as much as $2 trillion throughout the world."

Deloitte, October 2005

If the fundamental principle of AML monitoring is knowing your customer, then the execution of the process must encapsulate the many different types of customers your organization has, the ways customer accounts are established, the applications that manage customer touchpoints, and the different roles that individuals may play with respect to account activity.

For example, consider that, in the banking industry, there are many different types of accounts:

- Individual accounts, for accounts opened by individual people,
- Non-resident Alien accounts, opened for non-U.S. citizens originating from other countries,
- Commercial accounts for business entities, which may include numerous representatives who may act on behalf on the entity,
- Trust accounts, with trustees and others who may act on behalf of the trust,
- Foreign commercial accounts for business entities established outside of the U.S.,
- Personal investment corporations, and
- Institutional accounts (or other hedge or investment funds).

Not only that, each banking organization may have multiple customer touchpoints, with supporting applications developed in isolation of other organizational lines of business. Consider that a bank may have processes to open new customer accounts at a branch location, over the telephone or via the Internet. In addition, transactions and communication between the bank and the customer may take place via traditional telephony, mobile devices, branches, ATMs, wire transfers and through the Internet. Yet even these channels may eventually feed into a single repository — their disparity introduces some complexity into the ability to not just document who your customers are, but also to connect the many different types of transactions that might be considered suspicious activity.

One more complicating factor is the fact that individuals may play different roles in association with more than one account. The same individual whose name is associated with an individual account may also act as an advisor on a Trust account as well as be a partner of a corporation with a corporate account. To effectively monitor any suspicious activity, it is not sufficient to look at account activity, but rather to monitor the transactions performed by individuals in association with any of their related accounts.

Identifying Information

In order to effectively monitor transactions for the purpose of identifying possible money laundering, an organization must integrate identity management into all of its applications. This incorporates the management of identifying information associated with any individual, such as name, date of birth, any number of addresses (e.g., residence, business, residence of next of kin), telephone numbers, as well as some kind of identification number such as a social security or taxpayer identification number, or a passport number (for non-U.S. citizens). In some cases, even more information would be requested, such as the customer's net worth, annual income, occupation, employer and the customer's home country. Not only is this information used to verify the identity of a person, but it also becomes a key part of an ongoing process for tracking that person's activities.

To be even more thorough, recognize that suspicious activity may transcend an account and even customer boundaries. Suspicious activities may involve purchase and sale of real property, vehicles, vessels, aircraft, jewelry, equipment or other collectibles. The need to establish identifying information is therefore not limited to people, but may include the types of transactions that may take place and the objects involved in those transactions. Consequently, to really be able to know your customers, your application systems must be able to collect and use the right kinds of data to uniquely identify each customer within the context of the transactions in which that customer is involved.

The Objective: Identity Synchronization

Verification of the identity of a customer is not limited to collecting identifying information. The evolution of enterprise application systems introduces challenges to the ability for any organization to synchronize its customer views, which is the core capability for tracking customer behavior. The organic nature by which corporate data systems have evolved — largely designed to support tactical line-of-business needs, but not organized around a fundamental information architecture — has led to an environment where communication across applications is difficult, which in turn enables the (inadvertent or deliberate) obfuscation of identity. In other words, because the different interfaces and systems are unable to talk to each other, it is possible for a series of transactions to be performed that in isolation are innocent enough, but might otherwise be flags or indicators of suspicious behavior. Consider this: According to the Paris-based money-laundering watchdog Financial Action Task Force (FATF):[1]

"U.S. law enforcement has observed the following trends regarding wire transfers in terrorist financing investigations: (1) using "nominees" to provide clean names to terrorist financing transactions or accounts; (2) using front companies; (3) using multiple financial institutions; and (4) avoiding mainstream financial institutions."

Second, by virtue of the distributed nature of the application architecture, it is unlikely that there is a consolidated information architecture. This is manifested in a variety of data models and data stores, each with different representations of identifying information. One system may segregate given and family names, while others may have largely unstructured text that captures account names. Constrained data models may have encouraged data entry staff to shoehorn additional data into inappropriate fields, such as putting telephone numbers in the address line 2 field. Different applications may capture different addresses — the business account has the business address, but the individual account has the residential address.

Third, there are many opportunities for variation of representation to be introduced into a data environment. Despite the fact that each person is presumably assigned a name at birth, any single person might have many different name representations. Consider this example: a person may have established different accounts with the same organization using variations on his name. Robert Smith may be associated with one account as "Bob Smith" and at the same time may have taken out a home equity loan using his full name, "Robert Alan Smith." The original mortgage may be established under a joint account with his wife ("Robert and Mary Smith"), while the savings account he has held since childhood might be labeled "Bobby Smith." Despite the numerous names that represent this single person, it is still incumbent upon the organization to monitor whether Robert's transactions across all his affiliations indicate suspicious behavior.

1 *Third Mutual Evaluation Report on Anti-Money Laundering and Combating the Financing of Terrorism*, www.fatf-gafi.org/dataoecd/44/9/37101772.pdf

Indicators of Suspicious Activity

"46% of multinational banks do not have the capability to monitor a single customer's transaction and account status across several different countries."

KPMG Global Anti-Money Laundering Survey

As part of AML compliance, companies are required to report suspicious activity when it has been recognized, even though the parameters of what constitutes "suspicious behavior" may be deliberately vague to encourage best practices in review and analysis. However, clearly there are sentinel events or actions that, when coupled with communication or behavior patterns, would suggest further investigation. Many money laundering activities involve three steps: placement of unlawful money into the financial system, separating the money from its criminal activity, and using multiple transactions to provide the appearance of legality. Some examples might include:

- In the money transfer industry, a customer purchases money orders just below the reporting threshold with an apparent intention to avoid reporting;
- In the insurance industry, the transfer of a benefit on one customer's behalf to an individual with no apparent relation;
- In the real estate industry, a Politically Exposed Person (PEP) is involved in the purchase of numerous properties;
- In the banking industry, a customer engages in multiple transactions involving cash or cash equivalents whose amounts fall below the reporting threshold but whose sums exceed that threshold;
- A casino is requested to perform a wire transfer of funds whose source is cashed-in chips;
- In the securities industry, a customer executes numerous transactions involving certain kinds of security products traditionally employed in money laundering such as penny stocks or bearer bonds;
- In any industry, a customer decides not to proceed with a transaction when asked for identifying information.

Alternatively, it also includes transactions involving individuals who are known or suspected to be involved in terrorist activities. The U.S. Treasury Department Office of Foreign Assets Control (OFAC) maintains a list of countries, individuals and other kinds of organizations or entities considered to be involved in undesirable criminal activities (terrorism, drug trafficking, proliferation of weapons of mass destruction, etc.). Before transacting business with an individual, an organization should make sure that the individual does not appear on the OFAC list, nor is that person associated with any country or entity that appears on that list.

All of these examples reflect some similarity — an individual, potentially associated with some organization, attempts to execute a transaction involving some product, which correlates to some understood and defined "red flag" policy. Individual, organization, product, policy: all of these are named data objects that must be uniquely identifiable at many points throughout the enterprise, documented within a monitoring framework, and must be reviewable by an analyst as part of a monitoring program.

Anti-Money Laundering Compliance and Identity Resolution

Applications that support AML compliance must be designed to consider that at any time, any customer might be involved in some sequence of activities that could require scrutiny. The implication is that these kinds of applications must be able to monitor individuals, their actions (taken in sequences), their relationships to other individuals and their relationships to different kinds of organizational entities. There should be a process to classify the customers based on risk profiles, based on their demographics, income, and most particularly, their own employment or organizational positions. Individuals whose positions put them in proximity to financial activity require closer observation, as do those involved in cash-intensive business, such as pawnbrokers, high-price goods (leather, jewelry, car/boat/plane dealers), salvage business and even those associated with charities and other not-for-profits.

More importantly, AML applications should support algorithmic and statistical approaches for identifying and investigating "red flag" events, and this includes statistical modeling and customer profiling, as well as clustering individuals and activities, and even performing social network analysis based on matching and linkage models. Lastly, the process by which activities are monitored for compliance must itself be auditable, meaning that the business processes' support of AML can be reviewed by independent means.

From an information management standpoint, it should be clear that two significant requirements must be satisfied to most effectively deploy an anti-money laundering program:

1. Maintaining high-quality customer information; and
2. The ability to resolve variant representations of an individual to a unique customer (or organization, product or policy) record.

Fortunately, both of these requirements can be satisfied using a technique called Identity Resolution, which relies on approximate string matching technology. The Identity Resolution approach provides a set of methods to examine pairs of data values and develop a quantitative score reflecting the degree of similarity the two character strings share. Identity Resolution provides a means to evaluate whether two transactions were performed by the same customer, even though the names associated with the associated accounts were not exactly the same. Similarly, this approximate value matching process can help resolve addresses, telephone numbers and company names — among other objects that must be monitored.

Integrating Identity Resolution into your AML applications reduces the risks associated with both of our requirements. Embedding name searching and matching service at any customer touchpoint enables your applications to uniquely identify the individual and verify that person's demographic data, thereby maintaining high-quality information at each data entry point. Simultaneously, by uniquely identifying each individual each time contact is made, actions that are indicators of suspicious behavior can be logged, and sequences of events can be analyzed in real time to notify the designated compliance staff members and provide the data needed to file a suspicious activity report. Enabling linkage of records based on resolving variant attribute values also enables clustering, network connectivity, and the kinds of statistical analyses that analysts use to review seemingly non-connected events. And a predictably deterministic search and matching framework is robust under the scrutiny of audit and review.

The Challenge of Identity Searching and Matching

"Traditionally, thousands of hours have been devoted to the manual examination of large financial data sets. Data mining technologies have the capacity to streamline this process."

Tracking Dirty Proceeds:
Dept. of Criminal Justice
University of Central Florida

We have established that financial institutions must use names, company names and addresses to find or match information records. Yet the error and variation in such identity data is compounded by the volume of information records being searched and the need to perform searches in real time. The situation is compounded by the fact that it is typical in compliance and monitoring systems that, aside from the expected accidental errors and variations, criminals attempt to mask their activities by deliberately providing abnormal or extreme variations.

One set of approaches to resolving individual and corporate identities relies on phonetic compression techniques (such as Soundex or NY SIIS, which convert name strings into numeric codes representing phonemes) to assist in searching and matching; but these techniques on their own are limited by their inability to provide matches in ranked order, or to place name components in the right order (such as the difference between "Lee Kwok Ki" and "Kwok Ki Lee"), nor can they efficiently handle data from multiple countries/character sets/languages. Approaches that rely on data parsing and cleaning are far too sensitive to the need for detailed knowledge about the data to handle unpredictable data well, especially when dealing with foreign data. Solutions for non-Latin character sets (see Figure 1) that simply rely on transliteration to overcome the problem are dangerously simplistic because they do not adequately account for error in the original character set.

In fact, there is a plethora of opportunities for variation to creep into a name (and other identifying) information with the most common types of errors (Figure 2). The identity search problem is complex, and requires sophisticated tools and techniques that are reliable, accurate and deal with multi-country identity resolution to address the fundamental requirements of AML compliance.

Record 1	Record 2	Record 3
Peg Mc Cary	Margaret MacClary	Grietje McCllary
Abdulaziz A Rahman Al Sugair	Abd A Rhman Hammed Al-Shugair	عبدالعزيز عبدالرحمن الصقير
George Papadopoulos	Georgios Papadopoulos	Γεώργιος Παπαδόπουλος
Saito Kyoko	斉藤　京子	Kyouko Saitou
William Kwok	W. Kwok Ki Hoh	Mr. Billy H Kwok

Figure 1: Different types of character sets must be evaluated.

Name Variation

Names (and other identity data) suffer from unavoidable error and variation, which may include spelling, typing and phonetic error; synonyms and nicknames; Anglicization, and ethnic and foreign versions of names; initials, truncation and abbreviation; prefix and suffix variations; compound names; account names; missing words; extra words and word sequence variations as well as format, character and convention variations.

Common and Uncommon Words

The words used to label things are chosen from a vocabulary very different from meaningful language. There are no dictionaries, spell checkers or rules for the names of people, places, things or even addresses. The vocabulary in use for people's first names includes in excess of 2,500,000 words in the USA alone, yet as much as 80% of the population may have names from as few as 500 words. Accurate and high-performance name searching must perform for the uncommon names as well as for very common words. This is an extremely difficult challenge when a database of 100,000,000 people may contain 100,000 John Smiths, or Juan Rodriguez's or 1 Main Streets.

Variation or Error	Example
Sequence errors	Mark Douglas or Douglas Mark
Involuntary corrections	Browne – Brown
Concatenated names	Mary Anne, Maryanne
Nicknames and aliases	Chris – Christine, Christopher, Tina
Noise	Full stops, dashes, slashes, titles, apostrophes
Abbreviations	Wlm/William, Mfg/Manufacturing
Truncations	Credit Suisse First Bost
Prefix/suffix errors	MacDonald/McDonald/Donald
Spelling and typing errors	Porter, Beht
Transcription mistakes	Hannah, Hamah
Missing or extra tokens	George W Smith, George Smith, Smith
Foreign sourced data	Khader AL Ghamdi, Khadir A. AlGamdey
Unpredictable use of initials	John Alan Smith, J A Smith
Transposed characters	Johnson, Jhonson
Localization	Stanislav Milosovich – Stan Milo
Inaccurate dates	12/10/1915, 21/10/1951, 10121951, 00001951
Transliteration differences	Gang, Kang, Kwang
Phonetic errors	Graeme – Graham

Figure 2: Common errors and variations.

International Data

Most large identity databases contain data from multiple languages, countries and cultures that often have different structures, follow different parsing rules and have different variation characteristics. Also, if transliteration, Romanization, character set conversions and other such transformations are employed, a new class of error and variation is introduced.

Aliasing

Another wrinkle is the fact that it is possible that two people, companies and products might share the same name, while people, places, and things may be referred to using more than one name:

- People have maiden names and married names.
- People have aliases and professional names.
- Companies have registered names, trading names and division names.
- Places have several addresses, on two separate streets, old addresses, billing addresses, postal addresses etc.
- People and places can have names in more than one language.

The relationship between a data value and that object that the data value names is a many-to-many relationship, and indexing these multi-phased relations requires careful design in the majority of today's search applications. Searches that rely on a few values taken from a small number of fields are limited by the absence of semantic. Without the underlying context in which the value is used, the currency and accuracy of that data is called into question. Therefore, a reasonable identity-searching technique may require several keys or index entries pointing to the same identity to enable comprehensive resolution.

Identity Search and Match: What to Look For in a Tool

"The sobering truth is that poor data quality and data integration issues are often to blame for ineffective KYC or AML programs. And by the time the regulators have arrived, or criminal elements have laundered illegal gains, the damage has already been done."

George Marinos
National Data Quality Partner, PWC

For the purposes of AML compliance, consider integrating an identity search solution that overcomes both deliberate and inadvertent errors and variation in data, while maintaining system performance characteristics suitable to any necessary real-time constraints, yet is able to find valid candidate matches while limiting false matches. A reasonable product will provide:

- Intelligent and scalable algorithms, which, through the use of rich keys and search strategies, return all of the candidates an expert user would consider as being the same as the search criteria.
- Algorithms that are able to cope with real-world data, including data that is not formatted or cleansed or that contains missing, extra, truncated, out-of-order, nonstandard, or noise characters/words, initials, abbreviations, nicknames, numbers, codes and concatenations.
- Approaches that are enhanced through the use of a customizable rule base to incorporate the knowledge of the expert user, yet use a default population rule base to provide basic support out of the box.
- Functionality to support phonetic and orthographic correction functionality, to address spelling and typing errors.
- Intelligent matching routines that can be tuned to mimic the expert user making a choice as to which candidates are the correct matches. Such matching routines will incorporate all of the error and variation in the identities' attributes, as well as weighting the attributes as the user would.
- Algorithms that will work well regardless of the country of origin and language of the data, and must insulate the application developer from the differences between country and language.

identitysystems™

About the Author

David Loshin is the president of Knowledge Integrity, Inc., a consulting and development company focused on customized information management solutions, including information-quality solutions consulting, information-quality training and business-rules solutions. Loshin is the author of *Enterprise Knowledge Management — The Data Quality Approach and Business Intelligence — The Savvy Manager's Guide*, and is a frequent speaker on maximizing the value of information. www.knowledge-integrity.com

About Identity Systems

Identity Systems is a division of Nokia's (NYSE: NOK) Enterprise Solutions business group, is a global leader in identity searching and matching software, providing highly accurate and reliable solutions to search, find, match, screen and group identity data within computer systems and network databases. Identity Systems' solutions add data intelligence and quality to critical information-intensive systems in a range of vertical industries, including government, financial services, law enforcement and homeland security, healthcare and telecommunications. The company has built a client roster of more than 500 global organizations, including the U.S. Internal Revenue Service, Florida Department of Law Enforcement, GE Capital, Equifax, Experian, Bell Atlantic, Kaiser Permanente and Federal Express.

For More Information

Visit us on the Web at www.identitysystems.com or at the contacts below:

- USA, North and South America
  USASales@identitysystems.com
  Telephone (203) 698 2399
- UK and Europe
  UKSales@identitysystems.com
  Telephone +44 (118) 944 9688
- Australia and Asia
  AUSSales@identitysystems.com
  Telephone +61 (02) 9571 1300