ORACLE
Financial Services

A Guide to Effective Customer Screening

ORACLE WHITE PAPER | NOVEMBER 2017

Table of Contents

Introduction

This white paper is a guide for organizations choosing a system to meet their regulatory obligations, whilst minimizing the impact and cost on their business. Such a system should have the following key elements:

- Thorough data preparation
- Rigorous data standardization
- Sophisticated data and field matching
- Comprehensive case management

It should employ such techniques as sophisticated “fuzzy matching” algorithms and customizable workflows and match rules to reduce both risk and costs in meeting compliance obligations. The question compliance teams often raise is whether it is possible to reduce false positives without exposing their organizations to the undue risk of increased false negatives. With the right technology and strategies, this is indeed possible.

Screening for All Markets

Any company in the US must adhere to the Foreign Corrupt Practices Act (FCPA), the United (and) Strengthening America (by) Providing Appropriate Tools Required (to) Intercept (and) Obstruct Terrorism (USA PATRIOT) Act, and Export Administration Regulations (EARs). Companies operating in the UK must demonstrate compliance with the Proceeds of Crime Act 2002 (POCA) and Terrorism Act 2006.

Consequently, it's not just financial institutions that need to comply with anti-money laundering (AML) and counter terrorist financing (CFT) legislation. Mobile network operation third-party logistics (3PL) providers, trade exporters, and many others now need to undertake frequent and comprehensive screening to protect themselves from fines, brand damage, and potential criminal prosecution.

Achieving Regulatory Compliance

Legislation such as the USA PATRIOT Act, the European Union (EU)'s Third Money Laundering Directive, and other global anti-money laundering, know your customer (KYC), and counterterrorist financing regulations require firms to have effective systems and controls in place to ensure that entities such as criminals, terrorists, terrorist-financing organizations, and others are accurately identified and reported to the relevant authorities. Some regulations also require that firms identify and maintain a register of Politically Exposed Persons (PEPs) within their customer base.

While the majority of financial organizations have automate their KYC and customer due diligence (CDD) processes, many are dissatisfied with the cost and inflexibility associated with their current approaches. This, combined with a broader industry community now being exposed to the same legislation, is creating a demand for next generation watchlist screening solutions to maximize operational effectiveness whilst achieving optimal cost efficiency.

Figure 1. Government and intergovernmental groups, as well as industry advisors, have developed a vast range of regulations to fight the threat of money laundering.

Protecting Your Business from Risk

Rigorously screening your entire customer base on a frequent basis can alert you to the risk of being exposed to suspicious or sanctioned individuals and organizations – keeping your organization clean in both the regulatory and legislative sense.

Furthermore, purely from a business risk management perspective, knowing which of your customers, business partners, contractors, employees, or counterparties poses a heightened risk enables you to take the necessary

steps to protect your business and its brand and reputation in the market. Most current risk and compliance screening systems, however, fail by not providing the necessary capability in one or both of the following areas:

- Sophisticated data auditing, cleansing, and validation necessary to ensure that inaccurate, erroneous, and invalid customer data is removed prior to screening
- Advanced matching software algorithms (including powerful "fuzzy matching" logic) that are critical to accuracy and the minimization of false positives

Requirements for Effective Screening

Effective risk and compliance screening requires accurate data preparation, sophisticated matching, and comprehensive investigation, auditing and reporting processes. The following are key requirements for effective watchlist screening:

- **Identifying the highest risks.** Organizations that deal with potentially millions of customers need to prioritize their review activity against those that present the greatest threat.
- **Auditing and preparing source data ahead of screening.** With high levels of inaccuracies and inconsistencies often persisting in both customer data and watchlist sources, accurate data preparation is fundamental to achieving screening accuracy.
- **Accurately screening against a wide variety of risk sources.** These sources should include lists of sanctions, prohibitions, and PEPs and special interest persons; internal black lists and white lists; and lists of other internally sanctioned individuals and entities.
- **Screening of international data.** Because many of the world's major sanctions lists are scripted in Latin characters, accurate screening requires an ability to reliably match data across multiple scripts – including non-Latin alphabets such as Arabic, Chinese, and Cyrillic.
- **Fuzzy matching.** "Exact" and "inexact" name matching can be performed using powerful "fuzzy matching" algorithms for accurate identification – even when data is misspelled, incomplete, or sometimes missing.
- **Systematically screening across the whole organization.** Such screening is conducted at intervals appropriate to the risks faced by your business (usually daily) and in real-time when establishing new relationships with individuals and entities.
- **Integrating data from multiple sources.** Few organizations have all their data in one place and many store data in multiple file types and formats. An effective client screening system must be able to connect to multiple sources and automatically integrate different types of data from different systems.
- **Customizing workflows and match rules.** The system should allow for easy creation of multiple workflows and sets of match rules to accommodate a range of risk profiles.
- **Eliminating false positives.** Too many systems produce mountains of false positives every time screening is performed, requiring huge effort and wasting the time of compliance teams. Far greater accuracy than that provided in first generation screening systems is required to ensure that only genuine risk entities are identified.
- **Demonstrating enhanced due diligence.** Requires comprehensive audit trails, automated escalations, and extensive reporting tools with built-in case management system.
- **Eliminating repetitive review work.** Your system should remember your decisions, not produce the same matches over and over every time you screen your relationships. Once a user has reviewed a possible match, the system should not show the record again unless there is change in the underlying data.
- **Scaling according to your business.** Your system needs to process data volumes from very small (just a few thousand records) to extremely large (hundreds of millions of records) with the same rigor and accuracy. Smaller implementations should not be compromised by systems having limited capabilities. A scalable client screening system should be able to run on any size IT hardware unit – from a desktop computer to the data center machine.

Data Preparation is Critical

Data used in screening principally contains names, addresses, countries, and (in the case of individuals) dates of birth. For corporations or other entities, it also includes company registration and other details, plus data on key executives, stakeholders, and even beneficial owners. This data is likely to contain anomalies that will reduce the accuracy of your screening process, but it is also likely that the data you are screening against – the government and commercial lists – will also have anomalies or imperfections that will impact your screening results.

In addition, many organizations with a global footprint hold their data in international language scripts. This presents a problem when these organizations screen against sanctions and commercial watchlists, the majority of which are in Latin script and typically include limited translations of name data into Latin phonetic equivalents.

Increasingly, to effectively use content provided by both commercial and open-source list providers, organizations need to prepare data in a manner that goes beyond simple name cleansing. They need to include attributes such as nationality, country of residence, membership in certain regimes or political parties, close associates (otherwise known as "secondary identifiers"), and writing system used.

Extensive profiling and auditing of your data ahead of screening is therefore crucial. Techniques such as profiling, auditing, transforming, and analyzing text can be used to validate data against source systems. Often this will allow for removal of white spaces or questionable characters or split a single name field containing multiple attributes into a number of discrete fields, enabling data to be optimized to match rules.

Minimizing False Positives and False Negatives

False positives contribute to an excess of administrative work because they, like every other potential match, must be investigated. They require a significant amount of effort to sift through possible matches in search of supporting information or contradictory evidence. Simply put, false positives cost you money.

False negatives, however, can cost you your reputation. To claim that your organization failed to identify a sanctioned entity, criminal, or link to terrorist financing due to errors in underlying data is not considered a reasonable defense under government regulations or existing law.

False Positive: *A match produced by the system that, following subsequent investigation, turns out to be wrong.*
False Negative: *A genuine risk is missed by the system and is not flagged for attention. This can be seen as a weakness in the matching solution, but in practice, is often caused by poor quality of source data.*

One of the most effective (but often overlooked) methods of reducing false positives and regulatory risk is to ensure accurate preparation of data sources ahead of screening. This is because the accuracy of underlying data directly influences the results of any screening solution. The following are examples of how results can be skewed when poorly prepared data enters the screening system:

- Excessive numbers of false positives often occur when match rules cannot adequately process poor quality data. Invalid, incomplete, or inconsistent client data undermines most screening systems, resulting in mountains of repetitive false positives and screening criteria set so loose that the results are meaningless.
- Excessive numbers of false negatives can result from screening when inaccurate or poorly structured data enters the system. Data pollution is a fact of life in most organizations, yet it is intensely problematic for typical screening systems. For example, an entity name stored as "AIR-CON INC. T/A PRIME AIR-CONDITIONING SOLUTIONS" will not reliably trigger a match against "Air-Con Inc" in systems that are incapable of accurately parsing data and executing powerful fuzzy matching algorithms.

Once data preparation is complete, a careful balance needs to be struck in defining and applying the match rules for each risk profile being screened. Simply narrowing the screening criteria reduces the number of false positives; however, this increases the risk of a genuine threat evading detection. Therefore, quick and simple fine-tuning of highly granular match rules is needed to achieve an effective balance between minimizing both potential risk and the compliance team's workload.

Accurate screening also relies upon non-exact complex fuzzy matching. This is because criminal elements targeting organizations often transpose names, dates of birth, and other personal information to attempt to conceal their true identity.

Effective screening will differentiate between individuals and entities with common names, but will have discrete match rules available that can be activated. With the correct definition and application of rules to customer and list data sources, along with the use of secondary identifiers as part of the screening process, false positives can be reduced to a minimum without increasing risk.

Data Standardization

The standardization of data plays an important role in driving down false positives; it removes the need for inexact matching of data fields that can be better matched on an exact basis. Most data is, by its very nature, dynamic – new values arrive regularly. IT-intensive standardization of data introduces complexity and time lags into data preparation. Effective client screening systems should allow for "preflight" data standardization which can be accomplished by the compliance team itself, rather than having to involve strained IT resources and staff.

Understanding and Improving Data

Watchlist and customer data entries often include inconsistencies, inaccuracies, or are incomplete. Screening against data that is not fit for purpose reduces screening accuracy, increasing risks and costs. However, not all data quality challenges are as easy to remedy as simple field translations. Several other frequently-encountered challenges in screening are demonstrated by the example below.

Client ID	Client Name	Address1	Address2	Address3	Address4	Zip Code	Date of Birth
AD2329812	Mr M Jones (LVR) Trnsfr to CS Dept	Anchors Rest	38 Westward Avenue	Plymouth, Devon	PL17 3RF		43
VS38611	Mr J & Mrs A Probert	423 NE 2nd Terr	Fort Lauderdale	FL	U.S.A.	33201	13/02/1967 and 12/07/1966
DC182233	Mr P Adams T/A Heathcote Air Conditioning	PO Box 435	Brimingham	*** PRIORITY MAIL***		B1 3DS	
TZ350129	The Trust for Alan Jenkins Jnr	C/O Mr R Coffey	23 Bligh Cres	Rustington	Littlehampton	BN16 3ED	14/05/1987
CB278432	Deeter Von Furstenberg	21 Grand Drv		Grd Kayman		CI	

Figure 2. Ten common data errors are identified in this data sample.

Figure 2 demonstrates ten of the most common data errors, which increase data matching time and reduce accuracy.

1. **Poor spelling of name and address information.** A typing error such as "Micheal" or "Brimingham" can result in no-match decisions by a poorly configured screening application.
2. **Overfilling of name data.** Quite often, data gets appended into a name field to assist marketing or customer services departments, such as "Do Not Call" or "Deceased." Screening solutions often consider this additional information to be part of the name to be matched, potentially causing false negative results.
3. **Multiple names stored in a single field.** Joint account will include two or more names which need to be screened separately.

4. **Name information misfielded into address.** Sometimes the account name is an entity (such as a trust) which results in the first line of the address including the name of the head trustee. Without identifying the presence of a name in the address field, this individual will escape detection.
5. **Date-of-birth information in differing formats.** International differences in date conventions mean that both false positives and false negatives are possible.
6. **Entities and individuals mixed together.** This issue is similar to that of joint accounts; sometimes "Trading as" or "Care of" data is included in the name field, which again needs to be screened separately for complete protection.
7. **Non-standard name constructs.** You may encounter non-standard names, such as entities, within the name field.
8. **Poorly fielded address information.** Common issues include postal/zip codes and counties/states that are entered in generic address lines rather than in dedicated fields.
9. **Non-standard country information.** Combining data from multiple sources can result in standardization errors.
10. **Common false positive triggers.** Address information for a third party person in whose care the addressee is putting his mail can be a cause of false positives because it is mistaken for the actual person's address.

Working with International Data

A further layer of data preparation is required where customer data is held in multiple writing systems. A process is then required for converting non-Latin customer data into the Latin form as part of the data preparation. Techniques such as transliteration can be used to convert from one writing system to another using character-level rules, but more complex languages such as Arabic will require the use of transcription and variant matching logic to accurately identify all potential name equivalencies. One example of how these techniques are applied is given in Table 1.

Hangul	Latin	Variants
김	Kim	Gim, Khym, Kimm
이	Lee	Li, I
정	Jeong	Chung, Jung, Cheong
영	Yeong	Yung, Young
성	Seong	Sung
박	Park	Pak
경	Kyeong	Kyung, Gyeong, Gyung
수	Su	Soo
희	Hui	Hee, Hi
윤	Yun	Yoon
철	Cheol	Chul, Chol
호	Ho	Oh, O
조	Cho	Jo
진	Jin	Chin
석	Seok	Suk, Sok, Seog
현	Hyeon	Hyun
기	Ki	Gi, Kee, Gee
강	Kang	Gang
최	Choi	Choe
선	Seon	Sun

Table 1. Name variant dictionaries show, for example, the name Jeong can also be spelled Chung, Jung, and Cheong.

Table 1 shows how transcription and name variant dictionaries can be applied to customer data recorded in Hangul in preparedness for screening against watch-lists scripted in Latin characters. Names in Hangul first need to be converted to Latin, along with identification of all possible name variants.

Table 2 shows the standard name Kim has three potential variants. Each of these may appear as individual entries on a watch list and may, in fact, be the same person. The ability to apply the same techniques to secondary identifiers such as city, country or residence, and nationality is vital to increase screening accuracy and to reduce the amount of unnecessary review work by the compliance team. This work results from very high numbers of false positives.

Variant	Standard Name
Khym	Kim
Gim	Kim
Kimm	Kim
Pak	Park
Bhak	Park
Choe	Choi
Chwe	Choi
Chung	Jeong
Jung	Jeong
Chong	Jeong
Khang	Kang
Cho	Jo
Yoon	Yun
Chang	Jang
Yim	Lim
Im	Lim
Em	Lim
Hahn	Han
Ho	Oh
Shin	Sin
Sinn	Sin
Suh	Seo
Kwon	Kweon
An	Ahn

Table 2. Names in non-Latin languages translate to a wide range of Latin variants, just as data in other fields would.

Addressing data quality should be a fundamental part of all screening systems because deficiencies in data reduce matching accuracy and increase your risk. Compliance and IT staff are too often deluged with inappropriate system output, false positives, repetition, and reworking, which drains resources and hinders the business. An effective client screening system should have sophisticated data quality management capabilities to audit and profile incoming data – to cleanse, parse, restructure, and validate that data prior to screening. If done correctly, such intensive data preparation maximizes the quality of screening data, enabling the optimal use of phonetic and other data matching algorithms used in screening.

Conclusion

Regulations from government agencies worldwide designed to prevent financial crimes and terrorist activities have proliferated – vital protection indeed, but compliance with such regulations is challenging for organizations. This

white paper provides the reader with an overview of the key concepts to consider in maximizing the effectiveness of risk and compliance screening programs.

Key screening features and processes that can be used to achieve optimal cost efficiency were explained. Effective data preparation, whatever writing system is used to capture the original source data, is paramount in both minimizing the number of false positives and reducing the risk of false negatives. Techniques such as risk scoring and the use of secondary identifiers can also simplify identification of those individuals and entities contained within a PEP database that present the greatest threat to your organization – enabling you to take rapid action to eliminate unnecessary sources of risk. Finally, it is imperative that your chosen system has a suite of investigative and reporting tools (e.g. a comprehensive case management application) to eliminate unnecessary reviews. This will also ensure a clear audit trail, should it be necessary to demonstrate enhanced due diligence to the various regulatory authorities that operate within your jurisdiction.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

ORACLE

CONNECT WITH US

blogs.oracle.com/financialservices
facebook.com/oraclefs
twitter.com/oraclefs
oracle.com/financialservices

Integrated Cloud Applications & Platform Services

A Guide to Effective Customer Screening
November 2017