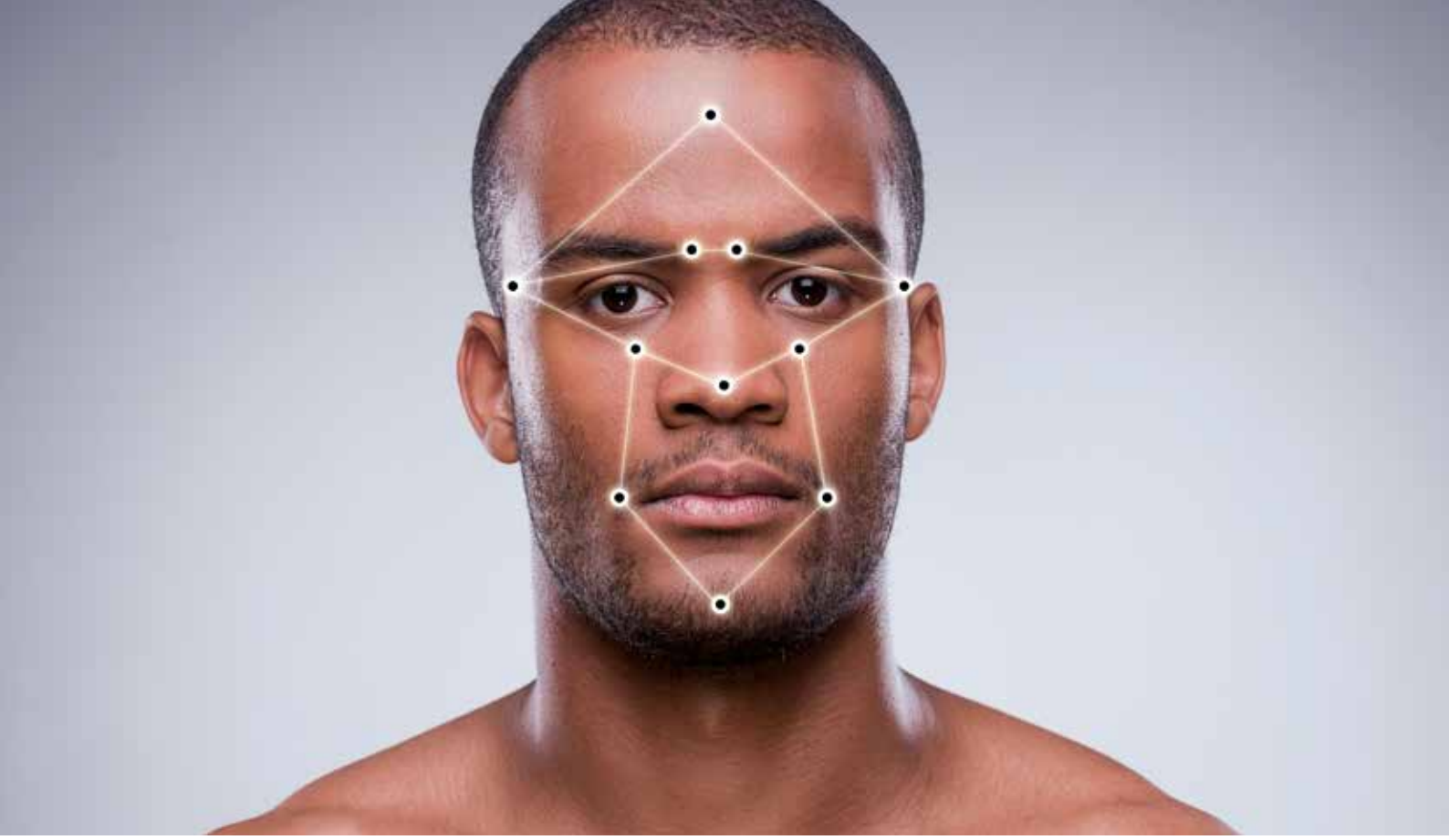
Challenges and trends in identity matching

Pascal Eber
Partner
Advisory & Consulting
Deloitte

Maxime Heckel
Director
Advisory & Consulting
Deloitte

Pascal Aerens
Head of Product Strategy
FircoSoft

Today we have come to accept the fact that our identity is checked for nearly every transaction we make. Among the checks performed by banks, shops, customs and airports, etc. a large number of them involve checking names against lists of sanctioned entities or high-risk persons. If you opened a bank account recently, your name will have been checked against these national and international lists before the account usage was authorised. The same applies whenever you take a plane, sign a mobile contract or even book a concert ticket.



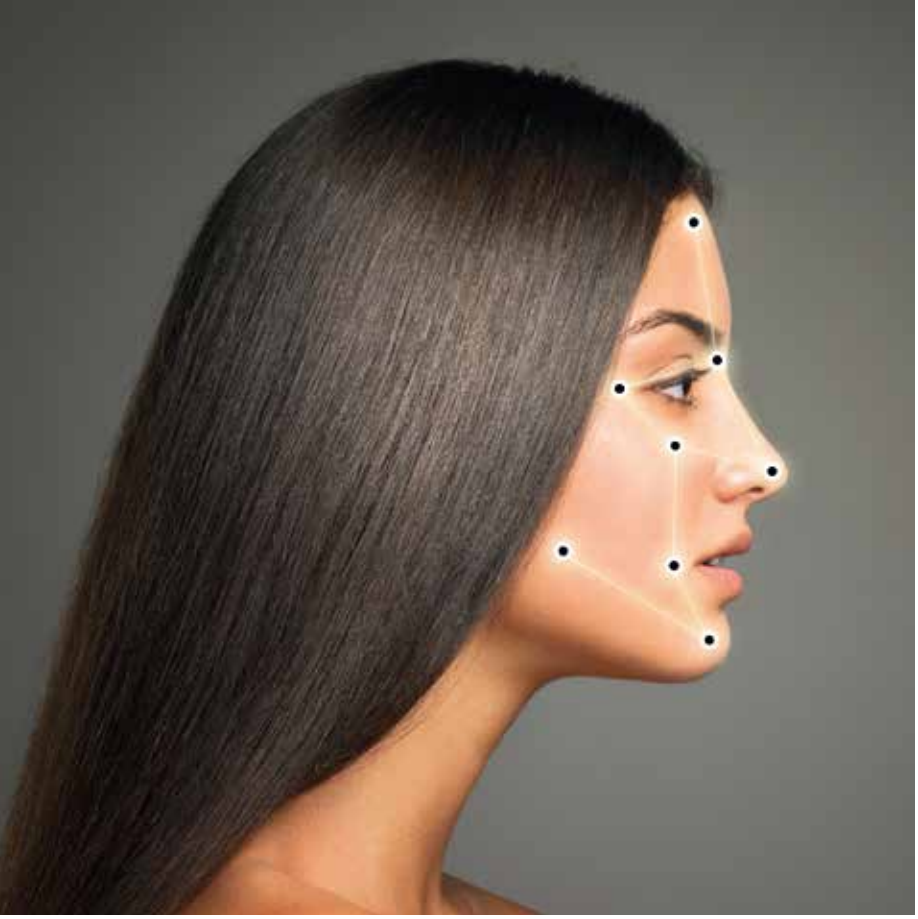
With regulations becoming increasingly strict, name screening has to be conducted in a growing number of sectors and against a growing number of lists. In the nineties, when electronic name screening was introduced (mainly for the purpose of enforcing the embargo on Cuba or blocking transactions for drug traffickers) the sanction lists barely contained a few hundred names. Following the terror attacks that marked the dawn of the 21st century, sanction programmes were boosted to detect terror funding and money laundering. As a result, there are now hundreds of lists worldwide, totalling millions of names, from 'political persons' to 'state-owned companies' to 'dual-use materials'. The sanctions on Libya during the Arab Spring, on certain African leaders, and more recently on Russia, clearly demonstrate that sanction programs now form an integral part of governments' foreign policies, replacing military action wherever possible. Within the current world context, it is clear that this trend will continue and probably accelerate in the future.

With regard to the practical implementation of sanctions, banks are de facto gatekeepers of the financial system. As such, they have been enrolled by authorities for their essential role in checking identities, either at account opening or when transactions are made. More recently, large corporates have understood that, even if they could delegate the screening of their relationships to the banks, their reputation is probably worth implementing their own preventive checks.

Challenges

The complexity of identity resolution, and particularly name matching, is often underestimated. A common opinion is that it is 'just' a case of finding a name on a list of names. Obviously, this is a huge oversimplification, as there are many technical challenges specific to this particular activity.

First, names are not just any string of text: names have a structure and typology that are influenced by many cultural and regional factors. As humans, we instantly recognise a French, English, Arabic or Chinese name, but teaching this cultural distinction to a machine is very hard. Simple algorithms that were developed to handle name matching rely on statistics to compute similarities and linear thresholds to trigger alerts or not. These simple approaches fail to produce correct results and often result in many 'false positives' (alerts that are triggered but not relevant). More sophisticated approaches use algorithms based on linguistics and machine-learning techniques to ensure no relevant hits are missed on the one hand, while controlling the operational costs of manual alert reviews on the other. Techniques based on linguistics also include translation and transliteration. 'Translation' is used to store variations that mean the same thing in different languages, such as *"Jean"*, *"John"* and *"Juan"* or *"Germany"*, *"Allemagne"* and *"Deutschland"*. 'Transliteration', on the other hand, is the ability to compare names in different alphabets (such as finding an Arabic name, written in the Cyrillic alphabet, on an English list). This capability is more useful than ever with the introduction of the latest sanctions resulting from the crisis in Ukraine.



The sheer quantity of data to be processed presents another challenge: comparing millions of listed names with millions of counterparties and/or transactions requires processing capacities that are several orders of magnitude higher than what was required ten years ago. During this period, we went from comparing a few thousand names with the U.S. sanction list (e.g. around 5,000 names for OFAC) to comparing full client databases (potentially tens of millions of counterparties) with Political Exposed Persons (PEP) lists that range anywhere between 1 and 3 million names. This means that the potential comparisons have grown from 10,000 x 5,000 to 10^7 x 10^6, excluding simple search techniques that come to mind first. Specialised search techniques, together with parallel processing, are needed to address these volumes.

In just a few years, Central Europe's geography has changed significantly, the political landscape in the Middle East has been dramatically upset and changes in Africa have affected millions of people, to name but a few

A third element to consider is that over time, screening systems have become completely embedded within the operational flows, making them a critical component. Today's screening solutions are 'screening hubs' that must be capable of processing checks and providing a reliable result in milliseconds. These solutions are implemented at the heart of the financial institutions' business functions and connected to all data streams, running on high-availability clustered platforms in order to share the computing load and offer sufficient scalability to handle volume peaks. Moreover, the global trend towards virtualisation provides yet another layer of flexibility to these architectures.

Trends

Innovations in name matching are focused on solving the major issues that will arise in the years to come: the operational cost of compliance, the volatility of the sanctions environment and the complexity of geographical sanctions.

The fact that processing power is available to do the number crunching for the screening process for large volumes of data has shifted research and development to the post-matching process. It is one thing to compute all possible matches, it is another to handle them. If the matching process yields a (very acceptable) 1% to 5% of detected records and the customer database is 10 million, it still means there will be at least 100,000 alerts to process manually. Therefore, the next algorithmic challenge is to reduce this to an acceptable number. Recent examples of multi-billion fines have demonstrated that regulators take sanction breaches seriously. This leaves financial institutions squeezed between the risk of huge penalties and the unbearable cost of manually investigating hundreds of thousands of alerts.

In order to solve this difficult equation, the current trend of innovation is focusing on assisting humans in the investigation process by partially or completely automating the decision process. Artificial Intelligence (AI) techniques, and particularly machine-learning, that have been evolving since the 1960s have acquired a degree of maturity that make them currently suitable for the use of false positive reduction and auto-resolution.

It is reasonable to predict that within five years more than half of the decisions on alerts will actually be taken by robots that have acquired their knowledge from the human investigators, letting the latter focus on complex cases requiring more experience.

The volatility of the global political situation has been a defining feature of recent history. In just a few years, Central Europe's geography has changed significantly, the political landscape in the Middle East has been dramatically upset and changes in Africa have affected millions of people, to name but a few. Consequently, sanctions and embargoes can be enacted and enforced in a matter of days, then lifted as quickly as they came if the situation improves. The level of reactivity required from screening solutions to allow financial institutions to be compliant with the latest regulations is dramatically higher than it was in the past. This means that the whole cycle of reference data updating, validation and implementation must be a robust, streamlined and fully automated process. Here again, automated agents can help by monitoring regulatory websites for changes, triggering automated processes that download, test and replicate reference data, warning users about changes and possible impacts and cross-checking multiple sources to assess the quality of the update. The reference data management of screening systems will likely move from the more or less static configuration used today to a large set of possible configurations—pre-set to respond to different crisis scenarios—that can be enabled on demand and activated instantly.

Lastly, the increasing complexity of the sanctions landscape is a major issue and solutions are implementing more and more sophisticated mechanisms to handle this. The first complexity that appeared was the move from fully sanctioned countries (e.g. Cuba) to targeted sanctions (i.e. specific entities, specific transactions and specific materials, etc.). This forced financial institutions to deploy complex rules in order to determine if a match on a name had to be considered or not. Recently, the situation has become even more complex: typically, the screening process only had to (efficiently) compare names in signaletic databases or transactions against sanctioned names present in a Watchlist. Any name not on the list is not sanctioned, and therefore not detected.

This changed with the last wave of sanctions against Russia: OFAC stipulates that entities where sanctioned persons have more than 50% direct or indirect ownership are also off-limits (even though these entities are not in the list).

This extension of sanctions to these indirect entities will require future solutions to incorporate ways to collect information from different online sources and implement 'big data' components that will retrieve, store and analyse data in order to detect relationships between listed and non-listed entities. This is clearly an exciting field of research, encompassing information gathering, data correlation and visualisation.

Conclusion

As sanctions are evolving from a minor political tool to an extension of foreign policy, underlying technologies have to adapt to address these challenges in our fast evolving world. The solutions for identity matching have only been in existence for 20 years and the changes made during this period have been remarkable. As regulations are expanding, we expect even more investment in research during the next five years. Banks have been alarmed by the immense cost of fines and have become extremely risk averse, but the operational costs associated with strict compliance policies are huge. Advances in technology are their only hope to balance these costs with their risk appetite.