BORDER
MANAGEMENT TODAY

WHAT'S HAPPENING AT THE EU BORDER?:
IBMATA HAS IT ALL

REFLECTIONS FROM PRAGUE:
A VERY SPECIAL GATHERING

DRIVE THROUGH BORDERS:
THE LATEST CHAPTER

GUARDING THE EU BORDER:
A VIEW FROM THE FRONT LINE

WE'RE GOING TO MIAMI
SEE YOU THERE?

By Farid Moussa,
Vice President of Strategy and
Public Sector, Babel Street

Closing the Risk-Confidence Gap:

Harnessing the Power of Open-Source Intelligence and AI-Powered Name Matching to Aid Decision-Making at the Border

Unsafe borders endanger citizens and stifle economic growth; failure simply isn't an option. Border security professionals need to make rapid decisions at points of entry to disrupt the flow of illegal goods and immigration, but dated technology and information overload hinders these efforts.

Experts have predicted that by 2025 over 463 exabytes of data will be generated each day globally. Not only are we seeing exponential growth in the volume of data, we are also experiencing growth in the veracity and the variety of data. With digital data growing exponentially, border security professionals need better and faster ways to separate the meaningful from the meaningless.

With the dark web, by design every user is attempting to obfuscate their identity, and bad actors are hiding much better. This presents a cat and mouse game wherein the cat must be smarter than the mouse, but the mouse is continually getting smarter.

Multilingual data poses additional challenges for detecting and preventing terrorism, transnational crime, cybercrime, intellectual property theft, and counterfeit goods — all of which threaten global and national security and economic vitality.

This is presenting a concerning Risk-Confidence Gap — a widening chasm between the escalating volume and variety of data that must be examined to obtain insight and

identify threats, and the resources available to analyze that data. Agencies have no way to confidently and efficiently parse the volume and variety of data — in different languages and scripts — for issues and threats.

Moreover, there's a need to minimize the high cognitive burden inherent in identity verification and remove repetitive and rote tasks.

As a result, we are increasingly seeing the need for open source intelligence (OSINT), enhanced with artificial intelligence (AI) and natural language processing (NLP) for improved threat intelligence decision-making for border screening.

Phonetic similarity
Jesus ↔ Heyzeus ↔ Haezoos

Transliteration spelling differences
Abdul Rasheed ↔ Abd al-Rashid

Nicknames
William ↔ Will ↔ Bill ↔ Billy

Missing spaces or hyphens
MaryEllen ↔ Mary Ellen ↔ Mary-Ellen

Titles and honorifics
Dr. ↔ Mr. ↔ Ph.D.

Truncated name components
Blankenship ↔ Blankensh

Gender
Jon Smith ↔ John Smith (but not Joan Smith)

Missing name components
Phillip Charles Carr ↔ Phillip Carr

Out-of-order name components
Diaz, Carlos Alfonzo ↔ Carlos Alfonzo Diaz

Initials
J. E. Smith ↔ James Earl Smith

Name split inconsistently across database fields
Rip · Van Winkle ↔ Rip Van · Winkle

Same name in multiple languages
Mao Zedong ↔ Мао Цзэдун ↔ 毛泽东 ↔ 毛澤東

Semantically similar names
PennyLuck Pharmaceuticals, Inc. ↔ PennyLuck Drugs, Co.

Semantically similar names across languages
Nippon Telegraph and Telephone Corporation ↔ 日本電信電話株式会社

Organizational aliases
Very Fine Groceries, LLC ↔ VFG

NAME MATCHING, USING AI, ANALYZES MULTIPLE CONTEXTUAL DATA POINTS ACROSS LANGUAGES TO ARRIVE AT MATCHES.

OPEN SOURCE INTELLIGENCE – UNLEASHING THE POWER OF UNCLASSIFIED DATA

While intelligence has traditionally been based on classified data, today, unclassified data is increasingly being used to provide context for other types of intelligence. Open source intelligence is collected by searching on topics or entities of interest that are publicly available on the internet at large.

In the past, OSINT has been seen as a "lesser form" of intelligence since it is unclassified. However, the view of the value of OSINT has changed dramatically over the past few years. SIGINT and HUMINT are the most expensive forms of intelligence, however OSINT is far more cost-effective. It's also becoming invaluable to support a key imperative: international communication and collaboration which is necessary to secure and reinforce borders for global trade and travel.

WHAT'S IN A NAME? CHALLENGES AND COMPLEXITIES ABOUND

Border control involves the scanning and matching of millions of names a day to secure borders and facilitate lawful international trade and travel, while enforcing laws and regulations.

Verifying identity has been an ongoing challenge for intelligence analysis due to the vast complexity of linguistics, spelling and cultural variances, human error, as well as human evasion.

It's tempting to think that name matching is like doing a keyword search, except the complexity of language makes it more challenging. New names are constantly created, with multiple spellings and no set of rules to encompass how names are formed. They are variable across languages, scripts, cultures, and ethnicities. Culturally-specific nicknames and aliases add to the complexity.

Some older name matching systems use a "brute force"

approach, generating all possible spelling variations to increase possible matches. This approach has weaknesses, such as dealing with not-yet-seen names and names with added or missing spaces.

Compounding this issue is the fact that more data than ever before is being published and posted in languages other than English. To this end, border security professionals must have linguistic capabilities that span hundreds of languages. Best-of-breed natural language processing capabilities run against the data while it is still in its native language. This minimizes the occurrence of analytic errors caused by inaccurate machine translations.

NLP recognizes the richness and variety of words and names in multiple languages and scripts, and their use across cultures. Using machine learning (ML) and linguistics algorithms, the technology simultaneously considers numerous types of name variations.

NLP and AI algorithms are employed to enhance datasets

for greater quality, usability, and completeness. Unstructured and relationship data are visualized through advanced link analysis, geographic heat maps, influential entity carousels, topic clouds, and patterns by time and day.

These advanced algorithms accurately score and prioritize critical entities within a relationship network while providing the citations from which an AI/ML-based decision is made. The technology has matured such that today, world-class entity matching technology measures over 100 features to calculate the similarity of entities across multiple languages.

TECHNOLOGY AS A FORCE-MULTIPLIER FOR MODERN BORDER MANAGEMENT

Modern border management cannot continue to rely on traditional tactics in today's vast digital landscape. We cannot "hire" our way out of this problem. Instead, it is imperative that we adopt technology to scale our efforts and free humans to solve the harder problems that machines cannot solve.

OSINT with AI and NLP is a critical force-multiplier; the technology can efficiently analyze massive volumes of data and distill actionable information, enabling analysis at speed and scale beyond human capacity. This is helping border security professionals overcome the Risk-Confidence Gap, while allowing humans to focus on the more difficult problems and/or vetting the results of AI.

Advances in technology are making it possible to streamline and scale cross-border screening prior to and during entry, while incorporating multilingual publicly available information and text analytics for enhanced decision making. With these technologies, border control professionals can put in place consistent, repeatable processes for automatically screening entities against public records, watchlists, social media presence, and other factors, that reduce the risk of human error and missed threats. As well, they can speed processing and reduce false positives, so as not to impede border crossing of legitimate trade and transit.

Farid Moussa is Babel Street's Vice President of Strategy and Public Sector. Prior to joining Babel Street, Farid retired from the National Security Agency as a member of the Senior Executive Service cadre. In his last position with the government, he led the Video, Image, Speech, and Text Analytics (VISTA) effort. Having guided language analysis teams for the counterterrorism target set in the post-9/11 era, he developed a deep appreciation for human language technology and its role in prosecuting the mission. Over the past 15 years, he drove the development of several VISTA service capabilities, empowering over 10,000 analysts within the U.S. Intelligence Community. Farid was a two-time keynote speaker at the Massachusetts Institute of Technology / Lincoln Lab biannual Human Language Technology Applications conference. He is a recipient of the Presidential Rank Award of Distinguished Rank.