A STRATEGY PAPER FROM CENTER FOR DIGITAL GOVERNMENT

RESOLVING IDENTITY:

THE IMPORTANCE OF WHO'S WHO AND THE SEARCH FOR THE PERFECT ENGINE

EVOLUTION OF IDENTIFICATION

The need to resolve individual identity is an age-old challenge. Identification documents have a long history — during World War I, American military officers carried identification papers at all times while on active service. These passport-sized identification books included a military portrait photo of the officer, basic biographical information and a signed attestation by a commanding officer or War Department representative. The portrait photographs were blind stamped with an offset official U.S. seal to provide further affirmation of the holder's identity.

During World War II, and in the years that followed, every American soldier had an official government military ID bearing a photograph, identification number and fingerprints of the individual. For the civilian population, the driver's license emerged as the method and means for accurately determining ID and was widely regarded as "the passport to America." New concerns about the security of the driver's license as a certified form of identification emerged in the wake of the 9/11 attacks and the review conducted by the federal 9/11 Commission.

"SECURE IDENTIFICATION SHOULD BEGIN IN THE UNITED STATES. THE FEDERAL GOVERNMENT SHOULD SET STANDARDS FOR THE ISSUANCE OF BIRTH CERTIFICATES AND SOURCES OF IDENTIFICATION, SUCH AS DRIVER'S LICENSES. FRAUD IN IDENTIFICATION DOCUMENTATION IS NO LONGER JUST A PROBLEM OF THEFT. AT MANY ENTRY POINTS TO VULNERABLE FACILITIES, INCLUDING GATES FOR BOARDING AIRCRAFT, SOURCES OF IDENTIFICATION ARE THE LAST OPPORTUNITY TO ENSURE THAT PEOPLE ARE WHO THEY SAY THEY ARE…"

— 9/11 Commission Report

"IT IS DIFFICULT TO THINK OF AN AREA OF DHS OPERATION WHERE A GREATER USE OF CUTTING-EDGE TECHNOLOGY WOULD NOT IMPROVE CAPABILITIES."

— Janet Napolitano, Director of Department of Homeland Security

As the significance of identification methods has evolved over the course of our nation's history, the use of paper IDs still begs a universal question that spans all levels of government and the private sector: "How can identity be verified with certainty, and what techniques and tools can be used to determine identity with as much accuracy as possible?"

In response to the 9/11 attacks, the federal government enacted a series of laws directed toward the goal of protecting the nation. Among them were the Patriot Act, the Aviation and Transportation Security Act, the Bioterrorism Act, the Enhanced Border Security and Visa Entry Reform Act, and the Homeland Security Act, which established the U.S. Department of Homeland Security (DHS).

In testimony before the House Committee on Homeland Security, DHS Secretary Janet Napolitano suggested that implementing common sense border and transportation security initiatives was one of her top priorities for DHS. In seeking to balance privacy issues with the enhanced documentation needed to better instill security confidence in the new identification standards that will be implemented as a result of Real ID — the 2005 federal law that imposes standards on state driver's licenses to be accepted by the federal government for "official purposes" — Napolitano said that "it is difficult to think of an area of DHS operation where a greater use of cutting-edge technology would not improve capabilities. Our border security efforts, port screening, transportation security, customs processes, immigration programs and preparedness and interoperability efforts could all benefit from a strong push to develop new technologies."

THE IMPORTANCE OF WHO'S WHO

Despite the advances in technologies and solutions, the concern that remains central to the objective of providing security to the nation is the question of "who is who?" The federal homeland security laws directed toward enhancing the security of the United States — and the federal government's confidence in the adequacy of such efforts — is, at the most basic level, grounded in the process of identity determination and the use of powerful databases to accurately determine individual identity to best evaluate threat and risk. Much of government's effort to protect the borders, minimize the risk to infrastructure and ensure the American public can travel safely, freely and efficiently — both domestically and abroad — is based on a complex system of computer-based data analyzers that search government records and provide information to help guide security personnel.

As one intelligence analyst employed by the National Security Agency (NSA) concluded, "Knowing the true identity of a subject — having the data available that proves you are who you say you are — is really the linchpin for a true homeland security strategy. When an individual presents him or herself at the border with identification documents, we need to know with reliability and accuracy that the person presenting the documents is in fact the individual reflected in the documents. We need to be able to do that quickly and with accuracy. Our security network is really based on the efficiency of that exchange, with the potential for unforgiveable consequences if we get it wrong."

Multiple questions tend to arise in the effort to ensure that the government does not "get it wrong" in protecting citizens. Among them are:

- ☑ How does technology reconcile disparate data sources to identify particular individuals with the accuracy required by government and demanded by the public?
- ☑ How can a high volume of records be efficiently analyzed by a technical solution to determine trends or make connections that a human analyst might not be able to recognize? In considering this, it is important to recognize that the volume of raw data is huge. The federal government, for example, maintains records on nearly 26 million U.S. veterans. Whether tax records or DMV records, the sheer volume of databases spread across the United States is daunting from an identity resolution perspective.
- ☑ How can a multiplicity of data stovepipes be reworked or filtered to make essential connections that help to determine identity?
- ☑ What are the practical uses of this technological capacity, and what are the means by which this process can be done quickly within minimal margins for error?

RESOLVING IDENTITY

Identity resolution, for the government and the private sector, is really a data-based intelligence process. While the private sector may be reluctant to embrace the concept of "intelligence gathering" to define its identity resolution efforts, the technological process really works on the same level. And its results, like most that involve technology, are only as good as the quality of the data available for survey. Identity resolution is the method by which organizations can establish connections between separate and independent data sources with the objective of identifying potential identity matches, as well as relationships that are not otherwise obvious to the human analyst. A robust identity resolution engine can analyze comprehensive batches of information relating to individuals or entities across a variety of databases, and then apply analytic programs to determine matches and the relationships that exist between the information contained in the data. Such search engines are regularly used to uncover fraud and help private businesses manage consumer data to both target commercial opportunities and eliminate redundancies in records.

IDENTITY RESOLUTION IS THE METHOD BY WHICH ORGANIZATIONS CAN ESTABLISH CONNECTIONS BETWEEN SEPARATE AND INDEPENDENT DATA SOURCES WITH THE OBJECTIVE OF IDENTIFYING POTENTIAL IDENTITY MATCHES, AS WELL AS RELATIONSHIPS THAT ARE NOT OBVIOUS TO THE HUMAN ANALYST.

In its simplest form, with regard to individuals, identity resolution works across different information silos such as employment records, tax data, commercial data and government watch lists. This is in order to reconcile various identification records to reach conclusions about the certainty of a particular individual's identity.

For a clearer picture, consider the following example. An individual named William Davis Smith has multiple record entries in various databases that are available for search. Records might have been generated under a given name, or in such variations as Bill Smith, W.D. Smith, W. Smith or even Davis Smith. By comparing consistencies between underlying factors such as home address, date of birth, Social Security number or other identifying criteria, an identity resolution engine can eliminate many records from the list of possible matches while confirming others that may not be readily apparent, or available, to the human analyst. Because of search engine technology use, attributes that were not immediately obvious — but that help to establish data connections, such as varying name spellings (or even aliases) — can be resolved to a particular individual because of commonalities in other aspects of the data check (i.e. bank accounts, addresses, phone records). Establishing connections between data that might not otherwise be identified by a human analyst offers a powerful and useful tool in definitively determining identity.

WHAT CAN IDENTITY RESOLUTION REALLY DO?

A well-designed identity resolution tool can instantly:

- ☑ perform criminal records checks based on information contained in multiple databases to accurately identify an individual with a criminal past;
- ☑ provide a better understanding of threat and risk based on a comprehensive picture of an individual's history;
- ☑ factor and sort multiple identities that may have been used by an individual over the course of many years and at dispersed locations; and
- ☑ meet federal Department of Homeland Security identification and documentation requirements.

MAKING CONNECTIONS

A country with unique geographical benefits recently embarked on a project to identify and intercept individuals who posed a threat before they entered the country.

The project is composed of a complex identity resolution system that analyzes various databases, including government watch lists, to report information to public safety personnel that even the most sophisticated biometric systems cannot provide. While biometric-based systems have the ability to definitively identify an individual, this identity resolution system goes further: it makes connections.

The system scans voluminous and seemingly unrelated databases to link information beyond the capability of human analysts. These connections not only confirm identity, but provide guidance to analysts to determine why public safety personnel might be interested in a particular individual. When the system identifies a subject of interest, a workflow application that is part of the system software routes the information to a variety of users (human and computer-based) for closer evaluation.

Although collaboration between agencies can often be a hurdle to overcome and was initially a challenge with the project, the ability of the organizations to access the system, conduct searches and receive results helped to mitigate early concerns, and has served as the basis for a solid public safety partnership.

"The value of the product is in the product itself," said one user, adding that "the flexibility of the system is one of its greatest benefits."

The project has been a success, because of participation and engagement by sponsoring government agencies in designing, deploying and fine tuning the pilot project, which was done in collaboration with a third-party vendor. The project has even earned praise from high-level leaders in the country.

UNIQUE IDENTITY: AN EXPERT'S VIEW

David Loshin, a private-business consultant who has studied identity resolution techniques and their application to both government and the business sector, defines identity resolution as "the ability to determine that two or more data representations can be resolved into one representation of a unique object." While public safety is the area where identity resolution has received the most media interest, Loshin recognizes the "growing need for the resolution of records associated with other kinds of data such as product names, product codes, object descriptions and reference data."

"Such considerations are not the only thing that is important, but serve as a prerequisite for all that follows," Loshin said. "We need to get it right. We are going to be engaged in this debate over effective identity resolution measures for a long time."

For homeland security, Loshin said that we need to turn the debate back around on the federal agencies and that we must be realistic. "It is all well and good that we have watch lists and systems in place, but the 20th time we stop the elderly man in line, we need to ask ourselves 'What are we doing?' and 'How can we do it better?' There is great potential use in the tools that are available, but better resolution tools will ensure better protection of individual rights instead of the subjugation of the masses."

According to Loshin, better tools are just one part of the equation. "In order to improve the process, we also need the agencies of government — and by that I mean both at the state level and the federal level — to work together."

EVOLVING TOWARD A BETTER SYSTEM AND THE NEED FOR SPEED

Identity resolution tools have reached a level of sophistication far beyond the Soundex system of the 20th century. Soundex, a technique that uses a phonetic algorithm for indexing names based on their sound when pronounced in English, is problematic given the demands of timeliness and accuracy required by modern users. The system can produce many versions of the same record, ignore matches and produce irrelevant or misleading results. In most applications, the technique also fails to properly rank order findings for a likely match. Problems have also been encountered in systems that use matchcodes for resolving records. The matchcode technique uses versions of key characters that, depending on the granularity, can produce inconsistent results that are difficult to accurately program for rank ordering.

In public safety, accuracy and timeliness are critical. In applications used by the federal government, the identities of air passengers need to be resolved from the time the passenger checks in at the gate to the time the plane departs. Given the realities of air travel and the convenience required by business and public passengers, there is no ability to transfer and process the export of records off-hours. The systems in place need to be reliable, accurate, timely and efficient. In achieving this goal, the systems need to maintain search integrity by filtering anomalous data entries to account for inadvertent or intentional record errors, or unique cultural name or spelling variations.

BEYOND PUBLIC SAFETY

While identity resolution systems are most frequently discussed in a public-safety context, such systems have been commercially used by the public and private sector to reconcile client and customer lists, as well as to target advertising and informational mailings. Increasingly, identity resolution tools are used to identify and eliminate fraud. The Texas Comptroller of Public Accounts deployed identity resolution technology to help the agency better administer tax collections. To more efficiently identify those who owed the state money, the agency contracted for a technology solution that would match tax files against other databases maintained by state and federal agencies. The solution generates approximately $70 million in annual revenue for the state.

Successful system models have also been developed internationally. The Office of State Revenue in New South Wales, Australia, launched an Unregistered Clients Project that identified more than 800 new taxpayers and recovered approximately $33 million in additional revenue. The project identified businesses that should have been registered for payroll taxes, but were not on record. While the primary goal was to identify businesses that had not properly registered for payroll tax purposes and to collect the revenue owed, deployment of the system resulted in efficiencies related to the reduction in the number of manual checks that had previously been required of investigators and staff.

IN THE SECURITY CONTEXT, IT HAS BEEN SAID MANY TIMES BY ELECTED OFFICIALS AND OTHER SECURITY PROFESSIONALS THAT "THOSE WHO WISH TO DO US HARM ONLY NEED TO SUCCEED ONCE, WHILE THOSE INVOLVED IN THE HOMELAND SECURITY ARENA NEED TO GET IT RIGHT EVERY SINGLE TIME, 100 PERCENT OF THE TIME."

A GLIMPSE OF THE FUTURE

Regardless of purpose, whether public safety or fraud detection, the success of identity resolution systems will ultimately be determined by the manner in which the balance between privacy and purpose, and speed and accuracy is addressed. In the security context, it has been said many times by elected officials and other security professionals that "those who wish to do us harm only need to succeed once, while those involved in the homeland security arena need to get it right every single time, 100 percent of the time." That need for accuracy is daunting when one considers all that needs to be factored into the dynamic, from privacy and the purpose of the inquiry, to the requirement for efficient speed of the inquiry transaction and the accuracy demanded from the result.

Systems that contain a comprehensive, unified and open platform for data integration can help balance these factors. These systems are designed to provide state-of-the-art technology in relation to identity resolution and have specific applications for government organizations with regard to law enforcement and fraud identification and prevention activities.

Search engines can support high-accuracy, high-volume searches of data to identify matches accurately and quickly, regardless of language, structure, format, location, duplication, omissions or transcription errors. Using complex algorithms and search methodologies, the systems locate identity records for large databases and discover the connections that are often buried within the records. The systems even compensate for data entry errors, intentional errors, and variations in international names and addresses so that identity determinations can be made with reliable certainty.

These systems are being used by government organizations throughout the world in a broad range of situations where names or addresses need to be searched, matched, grouped, screened or linked. There is a great diversity in the uses of such systems. Tax and Finance Departments have successfully used them to ensure taxpayer compliance. Law enforcement agencies locate stolen property; health and human service agencies track diseases and interpret pathology registries; and homeland security agencies secure borders and protect critical infrastructure.

THE SEARCH FOR THE PERFECT ENGINE

> *"THE ENTIRETY OF OUR ECONOMIC LIVELIHOOD OF THE 21ST CENTURY IS GOING TO TURN IN LARGE MEASURE UPON OUR ABILITY TO VERIFY IDENTITY."*
>
> — Secretary Michael Chertoff

The need to know "who's who" is an age-old dilemma, but it is one that can no longer be taken for granted. As technology evolves and becomes more and more sophisticated, so too have the efforts used by those who wish to avoid its detective capabilities. Meanwhile, privacy concerns and skepticism about the necessity for government to explore and resolve identity issues are likely to grow even more intense in the future. In addition, while system improvements will provide more advanced capabilities to mine a larger environment of databases, the need to develop more reliable and accurate resolution systems really serves as the foundation for the nation's homeland security structure.

If the identity question cannot be satisfactorily answered, then the nation's borders will not be safe. If a driver's license or passport used for identification does not accurately represent the identity of the individual who presents it, there is great risk in the uncertainty.

A state-of-the-art identity resolution system has the capability to override both intentional and inadvertent errors and variation in data, and can identify appropriate data matches while dismissing those that are false. Ideally, the system should explore connections between data through complex algorithms that an experienced intelligence analyst might explore given the dynamics of the search criteria. Such algorithms would be flexible enough to consider and work through the vagaries of data, with all of its consequent misspells and transcription errors, initials, abbreviations and numbers. The system would allow for customization, but provide a robust product and tool in its standard form that appropriately weighs certain attributes or factors that can help to determine identity. Most importantly, in this new era of homeland security and computer-based banking, the system should accommodate language differences and be applicable regardless of the country of origins to which the searches are applied and the data analyzed.

Toward the end of his tenure at DHS, Secretary Michael Chertoff said, "The entirety of our economic livelihood of the 21st century is going to turn in large measure upon our ability to verify identity."

In making this bold statement, Chertoff was referring to both the business and national security applications of the efforts to accurately determine identity. Robust sysytems are now available that provide a tool to help those charged with protecting the nation to better analyze and manage identity records, and help answer the often critical question of "who's who?"

INFORMATICA®
The Data Integration Company™

Informatica integrates scattered and isolated data assets so that government organizations can readily access all their information and trust its quality. Our platform discovers, distills, and delivers the unified knowledge agencies need to meet their requirements for effectiveness, transparency, and accountability.

CENTER FOR DIGITAL GOVERNMENT

The Center for Digital Government, a division of e.Republic, Inc., is a national research and advisory institute on information technology policies and best practices in state and local government. Through its diverse and dynamic programs and services, the Center provides public and private sector leaders with decision support, knowledge and opportunities to help them effectively incorporate new technologies in the 21st century.

www.centerdigitalgov.com

ACKNOWLEDGEMENTS:

Bill Howard, Senior Fellow, Center for Digital Education; Former Acting Chief of Staff, New York Governor; Former Chief of Staff, New York State Office of Homeland Security