



How to Verify Identities in a Complex Data Landscape

Government agencies must identify people and organizations accurately to perform their most essential functions — delivering benefits, preventing fraud, enforcing sanctions and more. But it can be a struggle to reconcile identities across multiple data sources, especially with inconsistent formats, language differences, and entities deliberately obscuring who they are.

Agencies often try to address these issues with cumbersome home-grown applications or software tools that aren't specific to the task. Those solutions can miss important connections, require multiple runs through the data, and introduce security issues. A better answer is entity resolution (ER) software that's purpose-built to pinpoint identities accurately, in real time.

This report looks at:

- The need for accurate identification of people, organizations, vehicles and vessels
- Why matching identities isn't as easy as it seems
- What to look for in ER software



Challenge: Verifying Identities From Varied Data

Definitively identifying people and organizations isn't as simple as it may seem. When agencies need a complete and accurate picture of a subject, they must compare multiple sources, each with its own issues.

- **Data about each of us is recorded in many places.** Your favorite online shopping site has your purchase history; credit bureaus track your debts. Government agencies also have multiple data sets for individuals, organizations, vehicles and vessels, which can involve billions of data points.
- **Agencies must bring those disparate sources together and reconcile the details.** But that's complicated by many factors. Michael may be recorded as Mike, and Mike and Mike, Jr. may share an address. Shell companies can obscure the ownership of an enterprise. Foreign addresses may refer to local landmarks rather than streets. And international records can use alphabets such Chinese or Arabic that require translation to match with English sources.
- **Processes must be explainable and transparent and protect privacy.** When data is used to approve or deny benefits or enforce regulations, agencies must demonstrate the process they used was accurate and fair. At the same time, much of the information may be personal or confidential and needs to be handled securely.

Solution: Purpose-built ER Software

Agencies need ER that addresses these complexities. Systems that aren't expressly designed for ER may stumble over differences in database configurations and fail to make the right connections. Dedicated ER software can meet government's unique identification needs.

- **Identify and track:** across multiple data sets, match entities across cultural and linguistic barriers, and identify ambiguous links for further review.
- **Real time:** ER that operates in real time helps identify cases so they don't slip through the cracks, whether that's resolving an eligibility question quickly so a beneficiary gets what they need, or identifying contraband before it enters a port.
- **In-house:** Software that stays in-house keeps data under agency control, protecting confidential and personally identifiable information. At the same time, it takes the mystery out of what factors were used to establish an identity, promoting transparency and explainability.

"Before you can start doing analytics, whether you're the U.S. Treasury administering the PPP loan program or you're CBP letting somebody into the country, you need to understand who you're dealing with. You want a 360-degree view of your citizen."

Will Layton

SVP Global Sales & Public Sector, Sensing



Best Practices in Intelligent Identification



Use an entity-centric system.

Many ER systems can only compare one data set against another, and may require running sequential reviews for each new set of data. An entity-centric system creates a persona that incorporates characteristics from multiple data sources and compares each new set to that persona, building a more complete picture.



Design for multilingual and multicultural data.

Trying to accommodate language and cultural differences on an ad hoc basis complicates the process and increases the risk of inaccurate and lost information. Interpreting disparate data formats should be integral to an ER system's design.



Account for ambiguity.

Inevitably, there will be instances where the connections between entities remain unclear. Is the John Smith who sells cars in London related to Jon Smith who sells cars in New York? Rather than discarding ambiguous data or confirming an uncertain link, good ER tags the ambiguity for further investigation.



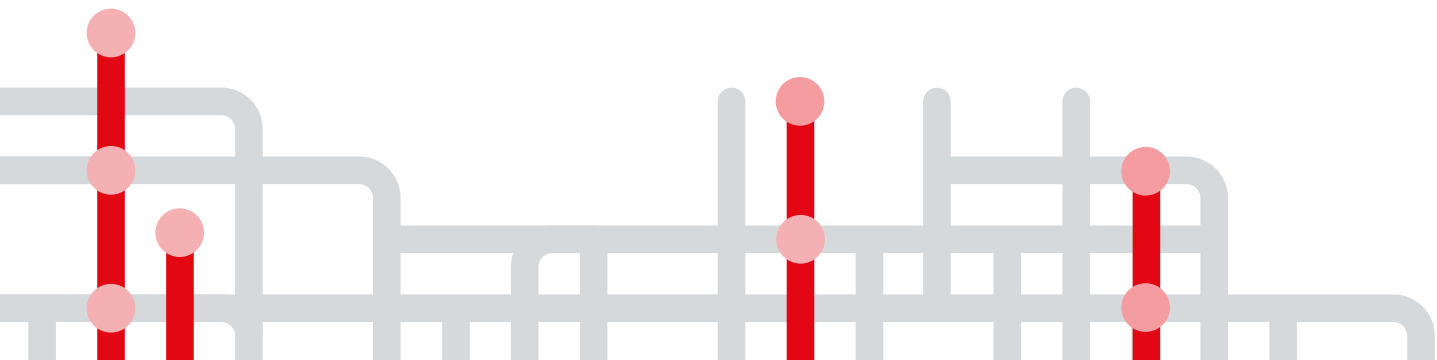
Preserve explainability and privacy.

ER systems make use of artificial intelligence. But sending agency data to a vendor's external AI introduces risks: The data may be retained by the AI and accidentally exposed, and the vendor's AI may operate as a "black box," blocking transparency. To avoid those risks, an ER should be under your control and do its work on your system, with clearly defined descriptors for decision-making.



Build for real-time processing.

For identity resolution to be useful, it must provide answers in real time. A system that requires multiple passes to determine connections is unwieldy and can miss important links. Good ER lets investigators find matches in seconds.





Use Cases: How ER Supports the Mission

- **The Paycheck Protection Program (PPP)**, designed to help small businesses during the COVID era, was abused by businesses that applied for multiple loans using slightly different names and addresses. By comparing data from multiple sources, ER software determined when two variants were the same person or company and prevented fraud.
- **U.S. Citizenship and Immigration Services** officials noticed that some applicants for entry and asylum used similar case stories, and needed to find connections between applicants and attorneys to detect potential fraud. A traditional matching approach didn't work, however. Entity-based ER software was able to trace the associations to uncover bad actors.
- **The Uyghur Forced Labor Prevention Act** prohibits importing of goods from the People's Republic of China that are produced using forced labor. But companies often use shell corporations and hidden relationships to obscure shipment origins. ER software compares third-party business data with import data to uncover the true source of products, so Customs and Border Protection can enforce the ban.

How Senzing Helps

Senzing is a leader in entity resolution software, with more than 250 years of combined experience building and deploying entity resolution systems. By delivering real-time, explainable results that identify hidden relationships, address ambiguities, and adapt to changing data, it allows agencies to make faster and more reliable decisions in areas from fraud detection to national security and regulatory compliance.

Learn more: <https://senzing.com/public-sector-entity-resolution/>