What Is Identity Intelligence? How It Works & Why It Matters

By: Jeff Jonas | AI Assisted | 100% Human Verified April 2026

Identity intelligence gives an organization a single, trusted answer to a question that touches every critical function in the business: *in what context do we know this person (or organization)? Are they a customer, an employee (or vendor), a fraudster — or are they all three?* It is what makes fraud detection, compliance, onboarding, marketing, and AI-driven decisions trustworthy.

"Identity intelligence" as defined in this article is distinct from the identity security products sometimes marketed under the same phrase, which focus on credential monitoring and threat detection. It is also distinct from *identity authentication*, which verifies who someone claims to be, as well as *identity and access management* (IAM), which governs who is allowed in and with what rights.

KEY TAKEAWAYS

- **The core capability:** Identity intelligence provides a continuously maintained, resolved view of who is who and who is related to whom. It delivers identities in context — whether called upon by AI agents, transactional workflows, other triggering events, or human questions — improving decision confidence in every case.
- **Intelligence through diversity:** Identity intelligence draws its strength from commingling data that organizations typically keep apart — customers and prospects, employees and vendors, good actors and bad actors, internal records and external watchlists. It is this diversity that connects the seemingly unrelated, revealing hidden connections and novel insights no single data source could surface on its own.
- **Mitigating identity risk:** Fragmented identity data spread across disconnected systems creates at least $500 billion in annual cost in the U.S. alone, spanning fraud, overpayments, compliance failures, wasted labor, regulatory fines, and more. This is the gap that identity intelligence infrastructure directly addresses.
- **Critical infrastructure for AI:** Agentic AI systems require identity intelligence as a foundational layer. Without it, autonomous agents risk accelerating and compounding an organization's existing identity risk.
- **Easily augments existing operations:** Identity intelligence is powered by an identity graph (a central index) that requires zero changes to existing systems; scales to a near limitless number of data sources and records; and answers queries in one shot, fast.

TABLE OF CONTENTS

What Is Identity Intelligence?

> **Identity intelligence** is the organizational capability that delivers identities in context: a single trusted view of *who is who* and *who is related to whom* across all eligible data sources. It is a single place where agentic AI and traditional workflows alike can look to accurately and confidently perform customer onboarding, customer service, compliance, fraud, risk, and other business-essential functions. Its power is proportional to the diversity of data commingled within it: internal data and external data, good guy data and bad guy data, master data and reference data (such as business registries, ownership hierarchies, and licensed parties). The more diverse the observational space, the more extraordinary the intelligence potential. The outcome: organizations make better decisions, faster. Sustaining this outcome is not a one-time project; it is an ongoing practice that spans technology, governance, and organizational commitment.

Ask yourself a simple but uncomfortable question: *where in your organization's infrastructure do you actually manage identity intelligence?* For most organizations, the honest answer is: Nowhere.

Despite massive investment in data lakes, data fabrics, and master data management (MDM), most organizations still cannot answer basic identity questions across their systems. The result: fragmented identity data spread across disconnected systems, a gap that costs at least $500 billion annually in the U.S. alone. Identity intelligence directly addresses this gap.

Most organizations do not have identity intelligence. What they have instead is a patchwork of fractional answers: different departments, different methods, different results. Many views of identity. Irreconcilable differences.

That gap between having *some* identity data and having *identity intelligence* is *identity risk* — and it is where billions of dollars in errors, fraud, compliance failures, missed opportunities, and more live.

The gap isn't a technology problem — it's a structural one. Identity data is spread across dozens of systems that were never designed to talk to each other. And the instinct to solve this by searching those systems one at a time (federated search) doesn't work. We'll explain exactly why below.

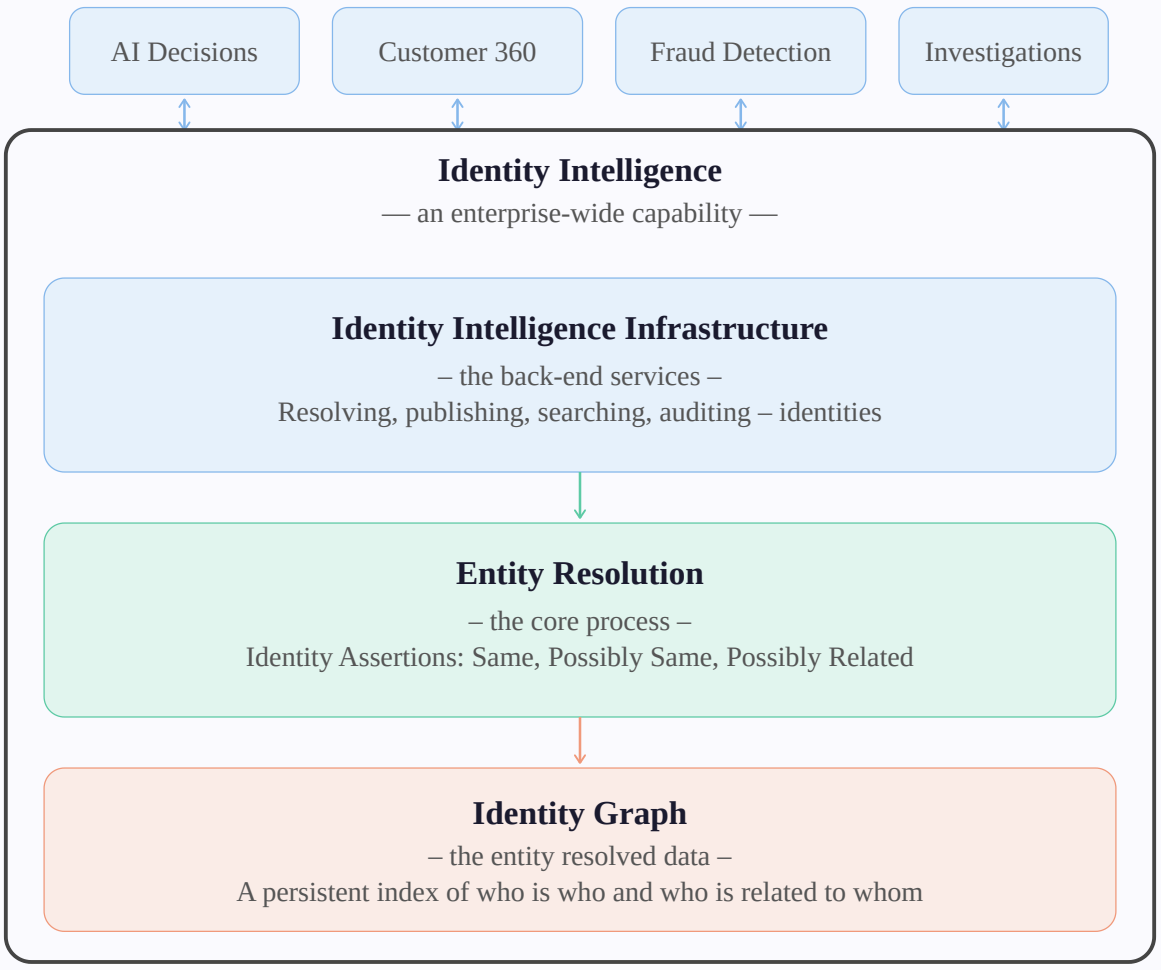
In the age of agentic AI, not knowing whether it's three customers with one account each or one customer with three accounts means AI decisions are flawed, and at machine speed, flawed decisions compound at scale. Most organizations have no strategy for addressing what's coming next: compounding identity risk.

How Does Identity Intelligence Work?

Every time an identity record is added, changed, or deleted anywhere in the enterprise (a new customer or vendor is onboarded, an account updated with a new address, an employee name change, a party removed from a watchlist), the relevant identity attributes are submitted to the identity intelligence infrastructure. This is the back-end service that resolves identities, publishes insights, replies to searches, and supports identity audits. At its core is entity resolution (ER): the process of determining who is who and who is related to whom, whether a person or an organization, and continuously managing the evolving identity graph with accuracy, currency, availability, and explainability.

Every record added, changed, or deleted is an *observation*: a new piece of evidence the system must evaluate against everything it already knows. With each observation, the system asks: *is this a known identity or a new one, and how does it relate to other known identities?*

The result is a living index — an identity graph: the entity-resolved view of who is who and who is related to whom, reflecting the current, resolved state of every identity across all eligible data sources. As would be expected with an index, the original data in each source system remains as-is, unaltered.



The data that lives in the identity resolved graph is purposeful, yet limited. Each record, submitted by a source system, will contain at minimum a reference to that source system and record ID, plus the personally identifiable information (PII) fields used for resolution — name, address, date of birth, identifiers, and similar. Most data sources will also include two other classes of metadata. The first: fields that help determine whether a given identity record is worth retrieving from the source system, such as whether a customer record is active or inactive, or whether an employee record represents a current, former, or terminated for cause employee. The second: fields that determine what can be shared with whom, such as classification levels and compartments in government or regulated environments.

The identity graph is not a system of record, not a data warehouse, and not a repository of transactions or events. It is a central index: a precision instrument designed to answer identity questions instantly, while leaving the full record exactly where it lives. Because the identity graph holds only the minimal fields needed for resolution plus lightweight metadata (not the full source record), it operates as a lightweight layer on top of existing systems. Production systems are not queried at query time; they are queried only when a user or agent decides a full record is warranted. This architecture delivers continuous identity intelligence without adding load to the systems that run the business.

This architecture requires zero changes to existing systems; scales to a near limitless number of data sources and records; and returns what's known about an identity fast. It answers in a single query what would otherwise require searching every system independently. Identity intelligence snaps into what you already have.

Architectural Qualities of a Central Index

limitless	~1	1	0
No. of data sources and records supported	Second to locate what's known about an identity	No. of queries it takes to locate what's known about the identity	No. of systems that must be first changed to support identity intelligence infrastructure

Identity intelligence snaps into what you have.

This capability is headless by design: accessible via APIs and function calls, built to be embedded in existing systems and summoned by autonomous agents, not operated through its own application or dashboard.

It is this identity graph that serves every downstream inquiry — whether via a human process, or an agentic AI fraud or compliance workflow, or a customer care chatbot. When someone or something asks *"is this person connected to a known fraudster?"* or *"is the ultimate beneficial owner (UBO) of this company a sanctioned party?"* or *"does this person already exist in our systems under a different name?"* — identity intelligence answers instantly and accurately. What it returns (subject to the requester's visibility rights) is the minimal content of the index plus attribution: reference to the relevant source systems and their specific records. The identity graph simply tells you where to look; assuming access rights are sufficient, the system of record can be queried for additional detail. Note: because entity resolution has been performed continuously as records arrive, the identity graph is always ready to reply with current information.

Beyond delivering identities in context, identity intelligence infrastructure is able to surface insights. One way to think about this is the notion of "data finds data, relevance finds you." With each new record ingested, entity resolution evaluates it against the existing identity graph — and when something noteworthy emerges, appropriate users or systems can be notified in real time. This means users don't need to know what question to ask — or when to ask it.

The revelations that may emerge with each new record are consequential:

- A new record resolves to an existing identity, with new contact information.
- A new record is deemed "possibly same" as a known fraudster.
- A relationship emerges between two previously disconnected identities, e.g., due to shared addresses.
- New evidence reveals that what was believed to be one identity is actually two, e.g., a junior and senior were conflated.

This last scenario: when a new record proves that a junior and a senior were previously resolved incorrectly, entity resolution with self-correcting properties will use the new observation to reverse the earlier assertion, ensuring up-to-the-second accuracy of the identity graph.

Why Does Identity Intelligence Matter?

Organizations are missing essential signals every day, only discovering them after the fact, when someone asks forensically: *how did that happen?* This is identity risk in action. Consider a few examples:

- A new customer is trying to onboard right now, but they are frighteningly similar (slightly different name, different passport, same date of birth) to a customer you off-boarded due to money laundering concerns.
- A manager in your fraud claims department shares an address (from their employment application) with three recently detected fraudsters.
- A small-account customer calls again, frustrated about consolidating their three loyalty cards — and that customer is a family member of one of your top ten revenue-producing customers.

How would you know? There are dozens, if not hundreds, of such scenarios playing out in any large organization at any given time — each one invisible until something goes wrong, and someone starts looking back through the data to understand how it was missed. Identity intelligence is what makes these connections visible in the moment — fast enough to do something about it while it's happening.

Research and real-world deployments consistently show that 10% to 30% of an organization's identity data is mismatched (over-matched and under-matched). Machine learning, AI workflows, and downstream business decisions are flawed to this degree. Every one of those errors has a cost — in dollars, in regulatory exposure, and in customer trust.

>$500B

Estimated annual cost of mismatched identity data in the U.S. — a conservative 12% share of the broader ~$4 trillion bad data problem. See basis →

Identity risk (the organizational exposure created by not knowing who is who and who is related to whom across an organization's data) is the invisible precondition for fraud, compliance failures, customer harm, and reputational damage. It compounds when AI systems act on unresolved data, and it grows with every system, data source, and autonomous agent an organization adds. The $500 billion figure above is a conservative estimate of direct, quantifiable costs. The broader strategic exposure — in competitive disadvantage, eroded customer trust, and compounding AI errors — is harder to quantify but no less real.

The impact is felt across the functions that touch identity data. Onboarding processes duplicate existing customers. Compliance teams can't establish a complete view of customer relationships. Risk models score the same entity differently depending on which system they query. Fraud teams miss connections between bad actors because records are fragmented. And AI systems — which depend on clean, resolved identity data to reason accurately — inherit every one of these errors.

> ***As noted by Gartner® in their research, "One of the big problems with data that comes from diverse sources for AI and analytics is that the identifiers for vital entities in the data tend to be inconsistent between data silos. In other words, the IDs for records from one data source are unlikely to match the IDs for the same records from other data sources. This prevents AI and analytics systems from combining or 'joining' this data in any meaningful way."***
>
> — Gartner, *Doing "Just Enough" Master Data Management for Analytics and AI*, Lyn Robison, 14 October 2025
>
> GARTNER is a trademark of Gartner, Inc. and/or its affiliates.

In an era where fraud including synthetic identities is proliferating at scale, the fundamental question organizations must answer is not just *is this person who they claim to be?* but *in what context do we know this person (or organization)?* Authentication alone cannot answer that question. Only identity intelligence — a resolved, continuously maintained view across all of an organization's data — can. Organizations that cannot answer the "who" question across their own systems are exposed to a class of risk that no single technology — firewall, fraud rule, or AI model — can fully address on its own.

Synthetic identities — derived from combinations of real and fake data, often using a real Social Security number paired with a fictitious name and date of birth — are particularly dangerous because there is no single victim to raise an alarm. They are one of the fastest-growing types of financial crime in the U.S., and they are purpose-built to exploit exactly the kind of fragmented identity data most organizations have today. Authentication cannot catch them — they were never real to begin with. Only identity intelligence, and its identity graph composed of diverse data sources, can surface the telltale signals: the same SSN appearing with different names, addresses that cluster suspiciously, identities that exist in one system but nowhere else.

Identity risk is not an IT problem. It is an enterprise problem, one that belongs on the agenda of boards, risk committees, and executive leadership. When a fraud goes undetected, when a compliance violation triggers a regulatory action, when a customer is misidentified and harmed — the consequences land on the organization, not the IT department. Whether the organization has identity intelligence infrastructure in place — and the risk exposure created by its absence — is a board-level concern.

The natural instinct when facing fragmented identity data is to search across systems one at a time — query the customer relationship management (CRM) system, then the fraud database, then the human resources (HR) system. This is federated search, and it's the typical approach. It doesn't work.

Why Does Federated Search Fail to Deliver Identity Intelligence?

The problem is structural. Each system of record was designed for its specific purpose — customer-facing systems built their own way, human resource systems fit for their purpose, fraud and investigative systems unique in design, as one would expect. This state of affairs presents four challenges that compound each other, rendering federated search a non-starter when trying to answer the question: *"what do we know about this person?"*

- **Lots of places to look** — customer management systems, loyalty systems, marketing databases, payroll, fraud reporting, data warehouses, email systems, and document management systems, to name a few. *Will anyone remember to search all of them?*
- **Data variations** — the same person may appear as "Liz Reston," "Elizabeth Smith," "Beth Smith-Reston," and "Liz Restn" across different systems. *If you are looking for "Beth Reston," will you take the time to try different variations? Will you think of all the variations? Can you even try all the variations? Or, where is the point of diminishing return?*
- **System search limitations** — each system has its own prescribed search interface, its own required fields, and its own rules about what can be queried and how. For example, many systems offer no way for the user to search by address or payment information — *are you going to ask IT to run a custom query against the backend?*

- **Recursive searching** — every new variation discovered demands going back to re-query every system already searched. If while searching system #11 it is revealed that "Liz Reston" is also known as "Elizabeth Smith," systems 1–10 should be re-searched using the new name variation. This recursive process is so time-consuming it's impractical for any organization to implement — meaning searches are declared done even though they are probably incomplete.

> ***Searching for a person across disconnected systems without Identity Intelligence is like searching a library without an index — you have to look through every floor, every aisle.***

Why Agentic AI Needs Identity Intelligence

Agentic AI systems are autonomous systems that reason, plan, and act on behalf of users, moving from proof-of-concept to production faster than most organizations are prepared for. These systems dynamically assemble workflows, spinning up tasks like reading from a PDF, running compliance checks, detecting fraud, and delivering next best action recommendations for customer interactions, without waiting for human instruction. They operate at a speed and scale no team of analysts can match.

At the center of many consequential agentic workflows lies an identity question: *who is this person? Is this the same company appearing under three different names across your systems? Is this transaction linked to a known bad actor? Is this customer already in your system under a different record?* These questions are answered by identity intelligence infrastructure.

Without identity intelligence as a foundational infrastructure layer, agentic AI systems are forced to query fragmented, unresolved data and reason on top of it. The errors don't disappear — they accelerate. If 10% to 30% of your organization's identity data is mismatched, what happens when autonomous agents are making decisions orders of magnitude faster? The error rate doesn't shrink. It compounds.

> ***Every agent workflow that involves a person or organization is only as trustworthy as the identity intelligence it's operating on.***

Organizations that deploy agentic AI without first building identity intelligence infrastructure are not just automating their operations — they are automating their errors.

This applies equally to decision intelligence platforms (DIP), the emerging class of software that models, orchestrates, and governs enterprise decisions. Whether rule-based, ML-driven, or agent-powered, every DIP that touches a person or an organization is a consumer of identity intelligence. The platform decides *what to do*. Identity intelligence tells it *who it's dealing with*.

And as autonomous agents increasingly act on behalf of humans — querying systems, making decisions, processing transactions — identity intelligence becomes even more essential: the organization must know not only which agent is acting, but who the human behind it is, and in what context that human is known.

What Are the Important Distinctions Between MDM and Identity Intelligence?

If identity intelligence is essential infrastructure, a natural question follows: *don't we already have this?* For many organizations, the answer they reach for is master data management. But MDM and identity intelligence are very different types of systems. Here are the four distinctions that matter most.

Distinction	Master Data Management (MDM)	Identity Intelligence
1. Scope	Owned, structured, internal data — active business entities	Owned and unowned data — including watchlists, investigative databases, and commercial reference data such as business registries and ownership hierarchies
2. Bad Data	Bad data is bad — eliminate errors at collection	Bad data is good — e.g., aliases, misspellings, and variations are valuable clues
3. Assertions	Firm facts — a "golden record" rarely reversed	Evolving predictions — constantly reevaluated as new observations arrive
4. Triggers	Event-triggered — activates during master data events	Everything is a trigger — every new observation is a candidate for attention

Distinction 1 — Scope

MDM is designed for data your organization owns and controls: structured master data that flows from internal systems. It deals with a known, bounded problem: the active business entities an organization generates: typically customers, products, vendors, and locations. One would not typically find marketing prospects, terminated employees, or watchlist subjects in an MDM system — yet it is exactly this type of diversity that identity intelligence must contend with. Identity intelligence spans the wider observational space: owned data plus data outside your direct control (watchlists, investigative databases, and so on) as well as commercial reference data such as business registries and ownership hierarchies.

Distinction 2 — Bad Data Is Bad vs. Bad Data Is Good

MDM seeks to eliminate errors at collection, expunging inconsistencies via a process known as data survivorship. Bad data is bad. Identity intelligence takes the opposite view: aliases, natural variability (Bob vs. Robert, St. vs. Street), old addresses, entry errors, disagreeing dates of birth (whether accidental or obfuscation tradecraft of criminals) are all potentially valuable clues. Bad data is good. Identity intelligence systems accumulate this diversity.

> ***In identity intelligence, bad data is good — it's often the only source capable of illuminating weak signal.***

Distinction 3 — Firm Facts vs. Evolving Predictions

MDM assertions are treated as firm facts, so reliable they are rarely reversed. Identity intelligence assertions are more like predictions: constantly postulated, reevaluated, and retroactively adjusted as new observations arrive. An MDM system manages toward a singular truth — a "golden record." An identity intelligence system manages a fully dynamic, ever-evolving identity graph where today's assertion may be revised by tomorrow's observation.

Distinction 4 — Event-Triggered vs. Everything Is a Trigger

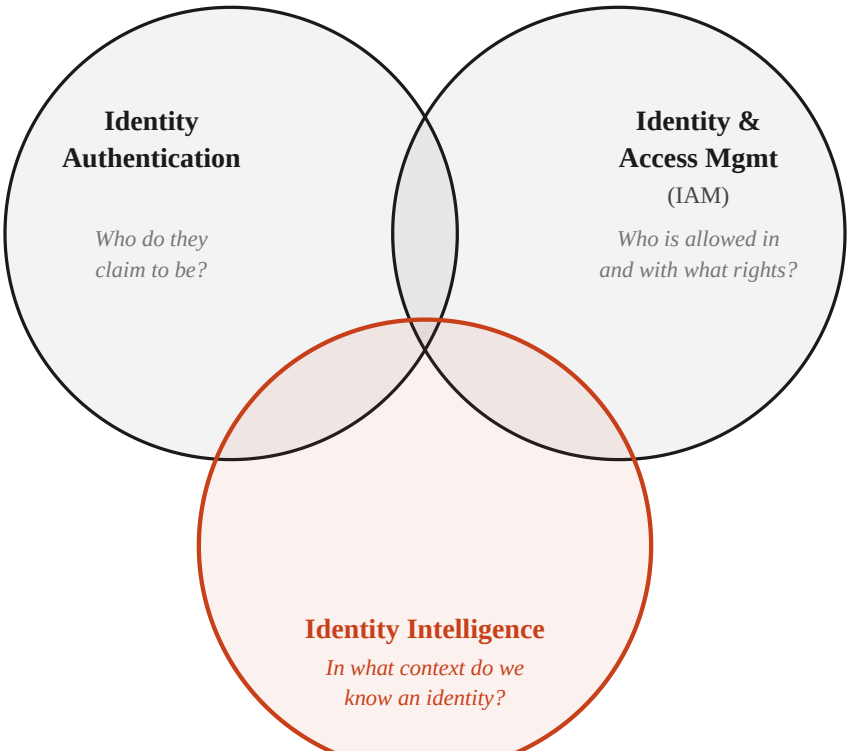
MDM processes are event-triggered: they activate during master data events like customer onboarding, account setup, or product creation. Identity intelligence treats every piece of data that enters the organization's observational space as a candidate for human (or agent) attention. With each new observation, the system must answer: *how does this relate to what I already know? Does this matter, and if so, to whom?* A change of address that aligns to an active investigation, an email address change that's associated with a VIP, a social service claimant who is a deceased person — each one is a potential signal.

MDM AND IDENTITY INTELLIGENCE: BETTER TOGETHER

These distinctions are not an argument against MDM. MDM excels at what it was built for: maintaining clean, authoritative operational records for the entities your business actively manages. Identity intelligence handles something beyond MDM — contextualizing data across a wider observational space, surfacing non-obvious connections, and treating inconsistency as signal rather than noise. But the relationship between the two depends on where an organization is in its journey.

For organizations that have already invested in MDM, identity intelligence is a natural expansion — MDM becomes one of the infrastructure's most trusted data sources, and the investment is amplified rather than replaced. For organizations that have not yet made that investment, identity intelligence may prove to be the more direct path to the outcome MDM was always trying to deliver: a single, trusted view of who is who. And for some, identity intelligence will reveal that the scope of the identity problem — spanning owned data, unowned data, good guy data, bad guy data, and data in disagreement — was always larger than MDM was designed to address.

What Is the Difference Between Identity Authentication, IAM, and Identity Intelligence?



The three identity disciplines — each essential, each distinct.

The word "identity" appears in several fields of technology, and that overlap could easily create confusion. Here's where the lines are.

Identity Intelligence vs. Identity Authentication

Authentication verifies that a person is who they claim to be at the point of a transaction — passwords, biometrics, SMS codes, multi-factor authentication, or knowledge-based challenges. Some practitioners further distinguish *identity verification* (the one-time confirmation of a real-world identity at onboarding — document checks, selfie liveness, government ID matching) from *identity authentication* (the ongoing confirmation that the person returning is the same one who originally enrolled). Both are point-in-time checks.

Identity intelligence answers a different question entirely: *in what context do we know them? Are they a customer, an employee (or vendor), a fraudster — or are they all three?* Where authentication operates on presented credentials, identity intelligence determines whether this is a new identity or a known one, and how they are related to other identities — across an enterprise-wide context.

The distinction matters in practice. An imposter walking into a bank with a fake ID and convincing the teller they are you is an identity authentication failure. So is a fraudster passing a selfie liveness check on a fintech app using a deepfake. A hospital merging two patient charts because "Robert Smith" and "Rob Smith" share a birthdate is an identity resolution failure — and a failure of identity intelligence. So is a digital healthcare concierge pulling the wrong patient's medication history, or a banking chatbot linking a caller to someone else's account because their names and dates of birth are similar. Knowing who is who and who is related to whom is a fundamentally different capability than verifying who someone claims to be.

Identity Intelligence vs. Identity and Access Management (IAM)

Identity and access management (IAM) concerns *who is allowed into a system and with what rights.* Identity intelligence answers a different question entirely: *who is who, and who is related to whom?* Both use the word "identity," but they address fundamentally different problems.

Bare Minimum vs. Advanced Capabilities: What to Look for in Identity Intelligence Infrastructure

Not all identity intelligence capabilities are created equal. Some are foundational: without them, the system cannot deliver on its core promise. Others are valuable force multipliers that separate decent implementations from great ones.

Bare Minimum: What Identity Intelligence Infrastructure Requires

These are the capabilities an identity intelligence infrastructure cannot function without:

- **Multi-source identity contributions** — the ability to combine and resolve identities across diverse, disparate data sources into a single view
- **Accurate entity resolution** — the ability to determine who is who and who is related to whom with accuracy that meets or beats human performance
- **Enterprise scale** — the ability to handle the volume, velocity, and variety of data the organization generates
- **Queryable identity context** — the ability to answer questions about an identity on demand, via APIs and function calls

- **Low-latency response** — the ability to return identity context fast enough for real-time decisioning, including autonomous AI agents operating within live transaction and workflow processing

Advanced Capabilities: What Sets Leaders Apart

These capabilities amplify the value of identity intelligence, and are increasingly expected in modern deployments [roughly in order of importance]:

- **Rapid data source onboarding** — the ability to integrate new data sources with minimal effort, in hours or days rather than weeks to months. For organizations with dozens or hundreds of eligible data sources, this is not just a technical convenience; it is a velocity advantage that translates directly to competitive advantage.
- **Agentic AI support** — the ability to make identity intelligence directly accessible to autonomous AI agents through a conversational interface, enabling agents to submit data for resolution in the identity graph, query for identity context, explain results, etc.
- **Audit-grade explainability** — the ability to explain any assertion: why two records resolved, or why records did not resolve, why related, and why not related — with sufficient detail to satisfy regulators and auditors when AI agents act on resolved identities
- **Relationship awareness** — the ability to surface connections between resolved identities
- **Real-time ingestion** — resolving and contextualizing identities as data arrives, rather than in batches
- **Operational simplicity** — the ability to deploy, configure, and operate the infrastructure without specialized entity resolution expertise, including native embeddability in agentic workflows.

- **Evolving identity notifications** — the ability to detect and publish changes in resolved identity context to any downstream system or process: whether fraud monitoring, updating a data warehouse, notifying an MDM system, continuous vetting, or keeping a graph database current. Every resolved change (a new match, a self-corrected false positive, a newly surfaced relationship) becomes an event that downstream consumers can subscribe to and act on.
- **Atomic record removal** — the ability to surgically remove a specific record and all its impact on the identity graph, essential for compliance with the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and similar data privacy regulations
- **Deployment flexibility** — the ability to run on any popular public cloud, your own private cloud, or on-premises bare metal environments for data sovereignty, security, and regulatory compliance
- **Application-level encryption** — the ability to encrypt identity data at the application layer, where the organization controls its own key creation, storage, and retrieval. This secures data even from the database administrator, reducing the risk of unintended disclosure regardless of who operates the underlying infrastructure. This is especially critical as cybersecurity breaches continue to accelerate: U.S. data breaches reached an all-time high in 2025, while open source supply chain attacks and the emergence of frontier AI models capable of discovering and exploiting software vulnerabilities en masse are expected to intensify the threat dramatically in 2026 and beyond. For use cases requiring discovery without disclosure (such as clean rooms, healthcare data environments, or cross-organizational resolution), application-level encryption is a critical enabler.

The bare minimum defines whether you have identity intelligence at all. The advanced capabilities define how far ahead of the curve your infrastructure is — and increasingly, how well-positioned your organization is to compete in an AI-driven world.

Case Studies: Identity Intelligence in the Real World

Identity intelligence infrastructure has been emerging for decades — quietly powering casinos, government agencies, and social services organizations long before most enterprises recognized it as a strategic capability. What's changed are the stakes. As organizations enter the agentic era, where AI systems make autonomous decisions at machine speed, identity intelligence has moved from a competitive advantage to an essential foundation.

Six Generations, One Team: How Entity Resolution Became Identity Intelligence

These five deployments trace the evolution of entity resolution technology across three decades — from 2nd generation to 6th. What started as one person's vision grew into a team that stayed together, accumulating over 300 years of collective innovation in entity resolution. Early systems proved the core concept: resolve identities across diverse data sources to surface what no single system could see alone. Each generation expanded the art of the possible: from batch processing to real-time resolution, from manual configuration to the autonomous onboarding of new data sources, from analyst-driven queries to agentic AI workflows. What began as entity resolution has matured into identity intelligence infrastructure, purpose-built for the agentic era.

Las Vegas Casinos — Non-Obvious Relationship Awareness (NORA) · 2nd Generation

In the 1990s, Las Vegas casinos faced a deceptively hard identity intelligence problem. Among the tens of millions of tourists visiting each year were a small number of known cheaters, banned players, and other subjects of interest — some barred by Nevada gaming regulators, others on the federal Office of Foreign Assets Control (OFAC) Specially Designated Nationals (SDN) watch list. Getting caught doing business with a banned individual could cost a casino its gaming license. But these individuals deliberately obfuscated their identities: slightly different names, transposed dates of birth, different aliases across different systems.

Our team at Systems Research & Development (SRD) built NORA — Non-Obvious Relationship Awareness — a 2nd generation entity resolution engine to solve this. The system combined identity data from across the widest imaginable array of systems across all casinos managed by the parent company — including every hotel reservation system, loyalty club, casino credit, accounts payable and vendor system, job applicant and employment system (current, former, and terminated for cause), arrest and investigative databases, and a wide range of internal and external watch lists. As a result, this near real-time identity graph contained ~15 million customers, 20,000+ employees, and 18 different watch lists.

The results, reported on this combined data, were striking. NORA alerts immediately identified 24 active VIP players as known cheaters, 23 players with relationships to prior arrests, 12 employees who were themselves active gamblers (in violation of policy at the time), 192 employees with possible vendor relationships, and 7 cases where an employee was the vendor.

The reaction from casino security leadership captured the broader point about identity intelligence. Jim Powers, Director of Corporate Security at Mirage Resorts, put it simply: *"Never underestimate what else this technology does."* The system had been built to catch bad actors at the door — but by placing all identities in the same identity graph, it enabled investigators to locate data and find hidden connections that would have been virtually impossible searching systems manually one by one.

Source: Jeff Jonas, IEEE Security & Privacy, "Threat and Fraud Intelligence, Las Vegas Style" (2006)

Cendant — Consolidated Consumer Database · 3rd Generation

Cendant Corporation was, at its peak, one of the largest consumer services companies in the world — a conglomerate whose brands touched nearly every dimension of American travel and homeownership. In hospitality, Cendant was the world's largest hotel franchisor, operating over 6,400 locations under brands including Days Inn, Ramada, Howard Johnson, Super 8, Travelodge, Knights Inn, and Wingate Inn. In real estate, its Century 21, Coldwell Banker, and ERA franchises were involved in one out of every four homes sold in the United States. In car rental, Avis and Budget made Cendant the second-largest rental car operation in the world. And in travel distribution, its Galileo Global Distribution System (GDS) held approximately 26% of worldwide computerized airline reservation bookings.

The identity intelligence challenge was staggering in both scale and diversity. Across these businesses, Cendant had over 5,000 data sources — hotel reservations, real estate transactions, rental car records, airline booking data, loyalty programs, and more, containing over 1 billion records. Our team built the Consolidated Consumer Database, a 3rd generation entity resolution system that resolved those billion-plus records into marketing profiles on approximately 100 million individuals. The result was a precision marketing capability unlike anything the industry had seen: the system could surface insights like *"Will he be more likely to respond to a promotion if we reach out 3 days before or 30 days before his next trip?"* This was identity intelligence applied not to fraud or compliance, but to understanding customer context and delivering timely sales offers, at a scale and specificity that no single brand's data could achieve alone.

Mike Kistner, Cendant, CTO: *"The Consolidated Consumer Database was my largest and most impactful project of my career."*

Alameda County Social Services — Much Faster Service, Way Less Fraud · 5th Generation

Alameda County Social Services Agency in Oakland, California faced a classic identity intelligence challenge: caseworkers were handling up to 600 cases each, manually searching across disconnected agency systems to piece together a complete picture of each client. The result was slow service delivery, missed benefit eligibility, and significant fraud exposure.

Working with IBM, the agency deployed IBM InfoSphere Identity Insight as the core of its Social Services Integrated Reporting System (SSIRS) — combining identity data across numerous social service and cross-agency systems to create a single, resolved view of every client. The result: caseworkers eliminated swivel-chair lookups entirely, the agency could proactively flag eligibility changes (such as a child turning 18 and qualifying for new benefits), and fraud was significantly reduced — at the front door before it landed in the fraud department. The independently audited outcome was a 631% ROI, with the initial investment recouped within two months of going live and ongoing savings of $24 million per year. The system won the 2011 Nucleus Research ROI Top Ten Award, the 2010 Computerworld Laureate, and IBM's Innovation Award for Outstanding Smarter Planet Solution.

Source: Computerworld Honors Laureate (2010); Nucleus Research ROI Award (2011); IBM Innovation Award for Outstanding Smarter Planet Solution.

ERIC — Improving Voter Roll Accuracy Across Member States · 6th Generation, Senzing

The Electronic Registration Information Center (ERIC) is a nonprofit, nonpartisan membership organization founded in 2012 to help states maintain accurate voter rolls. It has a bipartisan mission: help states remove ineligible voters from the rolls, correct out-of-date voter records, and help identify likely eligible citizens who haven't yet registered.

ERIC's identity intelligence infrastructure combines three diverse data sources: state voter registration data, state Department of Motor Vehicles (DMV) records, and Social Security Administration deceased persons data. By resolving identities across these sources, ERIC can determine who has moved out of state and forgotten to unregister, who has moved in-state and forgotten to update their record, who has moved into a state and hasn't yet registered, who is no longer among the living, and who may have voted illegally in multiple states. The system has identified almost 30 million in-state moves and updates previously undetected, flagged almost 14 million out-of-state moves, and identified over 659,000 deceased voters.

Source: Data provided by ERIC. See ericstates.org.

Fiserv — Identity Intelligence for Fraud · 6th Generation, Senzing

Fiserv (Nasdaq: FISV), an S&P 500 company and a leading global provider of payments and financial services technology, is among the world's leading banking and payment processing organizations. Supporting millions of merchant locations and thousands of financial institutions, the company operates at the intersection of banking and commerce — securely processing billions of financial transactions globally.

Leveraging a cloud-native data infrastructure on Azure, Fiserv employs Senzing's entity resolution solution at its core. This system assimilates data from hundreds of sources, resolving over 70 billion records into an identity graph comprising 10 billion records, ultimately resolving over 300 million unique entities. It enables online entity lookups in less than 20 milliseconds, meeting the demands of real-time payment processing.

Fiserv's fraud detection solutions gain enhanced insight into patterns previously undetectable at the transaction level. This approach establishes a robust foundation for fraud detection, risk modeling, and customer intelligence.

Source: Jay Duraisamy, SVP/CTO, Data Commerce Solutions, Fiserv — presentation at the 2025 Senzing Customer Summit.

All five examples share the same underlying architecture: diverse data sources — owned and unowned, clean and messy — resolved into a single trusted view of identity, enabling timely decisions that would otherwise be impossible. This is identity intelligence in practice.

Governance: Eligible Data, Privacy by Design, Data Sovereignty, Disclosure Control, and Oversight

The capability described above (and demonstrated in the case studies) is powerful. But power without guardrails is liability. As identity intelligence operates at enterprise scale, organizations face six foundational governance questions: *How do we determine which data is eligible for inclusion in our identity intelligence infrastructure? Where will the resulting identity graph — a central index — live, physically? Who has the right to access identity intelligence? Who has the right to see which records? How can every resolution decision be explained and defended? And how can every inquiry be traced and held accountable?* The sections below address each dimension.

Which Data Sources Are Eligible for Identity Intelligence?

The first governance question is also foundational: which data belongs in the identity intelligence infrastructure at all? The guiding principle is straightforward — the infrastructure should be bounded by what data it legitimately owns and is entitled to use. It is about enabling the left hand to know what the right hand holds. This is organizational self-awareness, and it is an obligation. It is the absence of this self-knowledge that enables unacceptable levels of fraud, poor customer experiences, and compliance failures. Organizations that cannot connect the dots across their own data are not protecting anyone. They are simply operating blind, inefficiently, and uncompetitively.

Not every identity-laden data source an organization possesses will contribute to the infrastructure. Some data sets are so highly protected (subject to attorney-client privilege, governed by strict confidentiality obligations, or carrying specific legal protections such as HIV status) that law and policy may deem them ineligible. Determining which data sources are eligible is not a technical decision. It is a legal, ethical, and governance decision that must be made deliberately and revisited as laws and policy evolve. These eligibility decisions will often involve considerations of core privacy principles — legal basis (including legitimate interest), purpose limitation, fairness and transparency, data minimization, and data subject rights that vary by jurisdiction and context.

Eligibility decisions also apply at the field level. Within an eligible data source, only the fields necessary for entity resolution — plus a limited amount of metadata — are needed. The guiding principle: use only the minimum data necessary for the specific purpose. Relevant metadata might include record recency (last date/time stamp updated), record status (active/inactive), and disclosure-controlling attributes (privilege or classification level) — just enough to determine whether fetching the full original source record is warranted, and whether the user, system, or AI agent asking the question has the privileges to access it.

A note for departmental owners concerned about control: identity intelligence is a mechanism for discovery and context, not a new system of record. It does not replicate, move, or take custody of your data. Your customer relationship management (CRM) platform, human resources (HR) platform, fraud system, and data warehouse remain exactly where they are, owned and operated by the teams responsible for them. The infrastructure simply builds a specialized index containing the minimal fields it needs — providing its users with a reference (pointer) back to your system should someone need more information. Your system retains its access controls and rules of visibility — control over who can see what. Think of it as enterprise-wide connective tissue: it makes every department's data more valuable — discoverable, contextual, and connected, without asking any department to give it up.

How Does Privacy by Design Apply to Identity Intelligence?

Several Privacy by Design (PbD) principles should be built into the identity intelligence infrastructure — and treated as requirements, not optional features.

Full Attribution

Every record submitted to the infrastructure must include a precise pointer back to its source system and source record. This full attribution is then retained — no data is lost or ever discarded, every data point is traceable, and thus any compliance question can be answered with a verifiable evidence chain.

Data Tethering

Adds, changes, and deletes in source systems must be published to the infrastructure. If someone is removed from a watchlist, deletion of personal data is requested under GDPR, CCPA, or similar regulations, or a conviction is sealed under a Clean Slate law, the infrastructure must be informed — and in a timely manner.

Single Subject Search

Identity intelligence enables an organization to answer a deceptively simple question: *what do we know about this person — across all of our data?* This single subject search capability is essential for compliance with regulations such as GDPR and CCPA, which grant data subjects the right to see everything an organization holds on them, request deletion of their data, or demand that their data not be used for analytics or profiling. Most organizations face the same problem this article describes: searching system by system, hoping nothing was missed, producing a response they cannot certify as complete. An identity graph makes this a single query — every record already resolved to the subject — enabling responses, deletions, and processing restrictions that are both complete and defensible.

False Negative Favoring

The entity resolution technology that powers identity intelligence should favor false negatives by default — only matching records when sure. It is far preferable from a civil liberties standpoint to miss a few matches than to make claims that are not true. Over-matching can deny a loan to a legitimate customer, block an innocent transaction, or misidentify a suspect. For use cases like marketing where over-matches are less consequential, possible matches can be collapsed into the base entity just-in-time when asked — without compromising the integrity of the underlying conservative identity graph.

Self-Correcting False Positives

If new evidence — a new record — reveals that two previously resolved entities are actually distinct (a junior versus a senior, for example), the system will need to disassemble the incorrect resolution and reassert the correct state, post haste.

> ***Self-correcting false positives are not just an accuracy feature. They are a civil liberties feature.***

Application-Level Encryption

When identity data is encrypted at the application layer — with keys controlled by the organization, not the database administrator — the risk of unintended disclosure drops significantly. Even if the database administrator's credentials are compromised, the underlying PII remains protected. This is also valuable in environments where identity intelligence infrastructure is operated by third parties, or where multiple organizations contribute data to a shared identity graph. Whether a clean room or other multi-party information sharing initiative, application-level encryption will play a key role in data protection.

For a deeper exploration of these Privacy by Design principles and their implementation in entity resolution, see: Privacy by Design (PbD) and Senzing.

Data Sovereignty: Where Does the Identity Graph Live?

Organizations must decide where this critical infrastructure — this persistent, ever-evolving identity graph — will reside, including the sovereignty of any replicas such as read-only copies. Strategically, an organization will want to maintain the freedom of action to decide where to deploy their identity intelligence infrastructure — whether on-premises, in their own private cloud, or via a managed SaaS service. For organizations with strict regulatory requirements, data residency mandates, or policies prohibiting data export to third parties (all of which can change), this architectural optionality is not just convenient. It is essential. In the agentic era, where autonomous processes are moving and transforming identity data at machine speed, maintaining complete control over where that data lives and who can access it becomes non-negotiable.

This also means that identity intelligence infrastructure is unlikely to be owned or operated by a frontier AI model provider. The identity graph contains an organization's most sensitive resolved identity data — the very data enterprises are unwilling to hand to external AI platforms over privacy, regulatory, and competitive concerns. AI agents may consume identity intelligence. They should not be the ones housing it. For identity intelligence to enable AI at enterprise scale, the infrastructure must be managed by the enterprise itself — running where the data lives, under the organization's own governance and control.

Who Gets to See What? Disclosure Control / Rules of Visibility

Just because data has been assembled into an identity graph does not mean everyone can see everything. Who is asking the question determines what can be presented to them. A fraud analyst or AI process may be entitled to see customer data and investigative data; a compliance officer or customer service representative may only be entitled to see customer data. Full attribution — data source pedigree — and metadata are used to inform the rules of visibility logic.

This becomes particularly important in a nuanced scenario: a data source that a given user is not privileged to see may have been the glue that brought records together — or added meaningful context that shaped how an identity was resolved. In such cases, the affected record must be redacted, and the identity graph partially disassembled so that the privileged data is removed from that user's view. The identity intelligence a user or agent receives must reflect only what they are entitled to know given their privileges.

How Is Every Identity Decision Explained and Defended?

Underpinning the entire governance framework is full explainability — the requirement that every resolution decision can be traced, explained, and defended. Whether a user or agent asks why two records were matched, not matched, or related, the system must be able to answer with a complete evidence chain. Without explainability, disclosure control becomes guesswork and governance becomes theater.

For organizations in regulated industries, "agentic AI" can sound like "uncontrolled AI" — autonomous processes making consequential decisions with no human accountability. Identity intelligence is designed to be the opposite: every high-stakes identity decision is traceable to a specific evidence chain — which records matched, why they matched, who asked, and when. Agentic entity resolution provides this intelligence. Authority, and accountability, remain with humans.

How Can Usage Be Held Accountable?

Identity intelligence is a powerful capability, and like any powerful capability, it can be abused. *What if an authorized user is searching not for a legitimate business purpose, but out of personal curiosity, or on behalf of a tabloid? What if they are searching for their daughter's new boyfriend, or looking up a famous actor? Who would ever know?* To address this, tamper-resistant audit logs can be used to record every inquiry made against the identity graph: which user asked, what they asked, when they asked it, and what was presented to them. Every inquiry — whether from a human user, an AI agent, or an automated system — should be written to a tamper-resistant audit log that can be reviewed by oversight committees or governance bodies.

The term "tamper-resistant" is deliberate. The goal is an audit log that even the database administrator cannot quietly alter — to cover their own tracks. One popular approach to implementing tamper-resistant audit logs is blockchain: by writing each audit event to an append-only, cryptographically chained ledger, the log becomes effectively immutable. Any attempt to alter or delete an entry breaks the chain and becomes immediately detectable. This kind of tamper-resistant audit infrastructure is not just a governance best practice — it is the foundation of accountability for any identity intelligence deployment.

About the Identity Intelligence Market Landscape

Senzing — the author's company — provides entity resolution infrastructure purpose-built for identity intelligence. Other companies with entity resolution capabilities that may service the identity intelligence market include: IBM Identity Insight, Quantexa, TiloRes, Verato, and Zingg. Companies whose MDM or integration platforms include entity resolution features — such as Boomi, Informatica, Profisee, Reltio, Semarchy, and Stibo Systems — may also evolve in this direction. And in the open-source space, Splink is a probabilistic record linkage library developed by the UK Ministry of Justice.

How to Get Started Today on Your Identity Intelligence Infrastructure

Before your next AI initiative goes into production, ask the hard question: *will the AI understand who it's dealing with?* If you can't point to a specific place in your infrastructure where identity intelligence is actively managed — trustable, explainable — then your AI initiatives are at risk. Not eventually. Now.

Legacy entity resolution solutions were not built for this moment — weeks of expert configuration per data source, accuracy drift between batch reloads, no support for autonomous agents. Agentic Entity Resolution is purpose-built for it, delivering continuously maintained identity intelligence — always current, trusted, accessible to every system and team that depends on knowing who is who.

Getting started does not require ripping out what you already have. Senzing can also be deployed as a sidecar alongside your current systems or as a parallel track powering new capabilities. Delivered as a composable library that runs where your data lives, it serves both agentic and traditional workloads from the same infrastructure. Learn more about upgrading in place →

Get Started: Senzing MCP Server. The Senzing MCP Server bridges AI assistants like Claude with the deep technical knowledge required to implement entity resolution — data mapping, software development kit (SDK) development, troubleshooting, and the full Senzing documentation library, all accessible through natural language.

Basis: >$500B Estimate

Harvard Business Review (Thomas C. Redman, 2016) estimated bad data costs the U.S. economy $3.1 trillion annually. Adjusted to 2026 dollars using U.S. Bureau of Labor Statistics CPI data (cumulative inflation 2016–2026: ~35%), that figure is approximately $4 trillion. Applying a conservative 12% attribution to identity data mismatch — records that are over-matched or under-matched — yields at least $500 billion. This is an illustrative estimate intended to frame the order of magnitude of the problem, not an independently audited figure.

→ Source: Harvard Business Review — Bad Data Costs the U.S. $3 Trillion Per Year (Thomas C. Redman, 2016)