GUIDE TO
Entity Resolution with AI-Native MDM

Tamr

At the core of every organization are the logical business entities that drive their operations, their strategies, and their success. From companies and suppliers to consumers, employees, students, and products, understanding and managing these critical entities – and gaining vital insights about them – is key to driving innovation, accelerating growth, and optimizing performance.

Resolving entities into single, unified views called "golden records" is a core component of master data management (MDM). Golden records foster greater collaboration across the business and equip everyone with the ability to speak the same language so they can agree on what the data – and metadata – is and is not. Golden records not only enable businesses to capitalize on the value of their internal structured data, but they can also use them to tap into the power of external and unstructured data, too.

In many data teams, "golden records" have a bad rap because their traditional MDM solution fell short on its promises. But golden records are turning their reputation around. In this ebook, you'll find out why.

Take investment firms as an example. When their fund managers are researching companies for potential investments, they need a holistic view of the target company, a consistent set of identifiers from data sources such as S&P, Capital IQ, or Pitchbook, and attributes such as LinkedIn employee count growth consolidated into a single, unified view—the golden record. Without these golden records, they lack context and insight into the business data that matters most to their investment criteria, hindering their ability to make informed investment decisions. Ultimately, this leads to missed investment opportunities, poor investment choices, and a significant negative impact on the firm's overall business performance.

To capitalize on the true value of their data, and use it to drive better decisions, businesses need to solve the hard problem of entity resolution. But what, exactly, is entity resolution? And what makes it so difficult?

What is Entity Resolution?

Entity resolution, also known as entity linkage or record matching, is a data management technique used to associate multiple disparate datasets into a logical entity such as a person, business, or product. Entity resolution addresses the challenge of reconciling records across and within datasets by detecting and matching records that are the same, despite differences in spelling, formatting, associated attributes, and other discrepancies. From there, entity resolution assigns a unique ID to these matched records to ensure that they are treated as one unique entity going forward.

For example, if a company sells the same product online via Amazon and Walmart, the product may have different names, descriptions, and unique IDs on each site's respective listing pages. Because the entities are not an exact match, analyzing total sales of the product is difficult because items that are actually the same appear to be different. The company could assign resources to manually scrape the product pages of these websites to track sales volume and price, but the resources would quickly find the task incredibly time-consuming and nearly impossible to do at scale.

Why is Entity Resolution So Hard?

Experts predict the world's data will top 181 zettabytes in 2025, and that data volume will continue to surge. To put these numbers into perspective, in 2013, the volume of data reached just nine zettabytes.

While accelerating volumes of data make entity resolution difficult, other factors come into play as well, including:

1
Heterogeneity of data quality.
Data from disparate sources are likely to use different formats for data points such as dates, addresses, or abbreviations. On top of that, misspellings, missing information, and even intentional manipulation are common for large quantities of data. All of these factors can contribute to duplicate data or data points that are difficult to connect across a multitude of sources.

2
The problem of scale.
Theoretically, if there are n records in total, then you need to compare n-squared pairs to reconcile the data. But because data is so diverse, it becomes almost impossible to write rules to match them. A human can often look at two records and determine if they refer to the same thing or not. But this manual reconciliation does not scale across all records, even for relatively small datasets.

3
Changing business context.
Each use case requires specific data sources and has different requirements around the precision and recall trade-offs. For example, a compliance analyst investigating financial crime would want to include more fuzzy matches whereas a patient matching use case would want to avoid linking two different patients as one. That's why it's essential to have multiple views of entities for different requirements.

4
Missing external context.
Sometimes the data you have just doesn't tell the full story. Common scenarios include company name changes, company acronyms, a person changing their work email, and a person changing their last name. Without external data, it can be impossible to know if two entities are the same.

Gaining a single, authoritative, accurate version of business entity data across multiple data sources and datasets is hard work. And as data continues its upward trajectory, businesses will continue to struggle with this challenge. Messy, incomplete, inconsistent data that's often out-of-date will remain spread across disparate systems and silos throughout the business causing different users to struggle in different ways.

Data engineers must manage immense volumes of data from a litany of disparate sources. They question the data's quality, but because they're armed with the wrong tools, they can't resolve and match data efficiently and effectively.

Data consumers need data to do their jobs. But they are never sure if they can trust it.

Data analysts are responsible for analyzing data to uncover insights, identify trends, and generate reports that will be used by others within the organization, but they're left with inaccurate analyses and flawed insights that potentially lead to poor strategic decisions.

Business end users, including departments such as portfolio management, procurement, marketing, and risk and compliance, consume data to make business decisions, develop strategies, or optimize operations. But they find that inaccurate data leads to inefficient processes. As a result, they spend significant time correcting errors and verifying data which not only wastes their time, but also slows down workflows.

Data stewards, tasked with maintaining data quality through the management of data standards and governance across the

organization, face overwhelming challenges caused by the volume of possible matches and potential anomalies. Consequently, they're drowning in a data deluge, which compromises data quality and hinders organizational decision-making.

While the challenges with entity resolution may seem insurmountable, there are ways to overcome them. Innovative, **AI-native MDM** solutions provide limitless opportunities to adapt and scale as data continues to evolve and grow both in size and complexity, enabling organizations to finally realize the promise of golden records.

10 Key Principles of AI-Native MDM

Humans can perform entity resolution well, but the problem is that they're slow. AI algorithms can do things fast and at scale, but they are too rigid to operate completely on their own. That's why the best solutions provide the best of both worlds when it comes to entity resolution: the power of AI and the expertise of human refinement.

Human refinement (aka, human intelligence) lends credence to the valuable role humans play to ensure golden records represent the best, most trustworthy version of key business entities. Human refinement strengthens AI and enhances its learning, helping the AI to become increasingly more accurate over time. Think of human refinement as a safety check, providing a critical layer of oversight and finesse that complements and validates the AI's capabilities.

In addition to human refinement, the best solutions embrace ten key principles, including:

1

Time-to-value: new data sources, both internal and external, are emerging all the time. The best solutions allow you to onboard new data sources quickly and easily so you can realize value faster. And they come with built-in integrations to external data sources to save you even more time.

2

Scalability: as your organization grows, so does your data volume. It's vital that the technology you choose is cloud-based and optimized for scalability so it can handle the largest volumes of data possible.

3

Machine learning: look for solutions that use a machine learning-first approach to entity resolution. Machine learning improves with more data, increasing automation and freeing up technical resources by up to 90%. Rules-only approaches have the opposite effect.

4

Accuracy: a person can usually determine if two records are the same entity, even when dealing with poor quality data and incomplete information. That's why AI-native MDM solutions should be at least as accurate as a person doing the work manually. But remember, a human should still verify the data matches AI produces to ensure the highest level of accuracy.

5

Flexibility: one size does not fit all. Depending on your use case, you might want to split up entities or roll them up to the same entity. Your technology should enable a flexible approach so you can manage different use cases in different ways, and easily adjust as your needs evolve.

6

Persistent ID: data and attributes change over time. As you reconcile different records into the same entity, it's important to maintain a unique, persistent ID in order to have a longitudinal view of the entity. The Tamr ID solves the challenge of data fragmentation. Learn more in this **post**.

7

Data lineage: as data records move in and out of a cluster because of data changes, it is important to keep track of the record-level data provenance. Additionally, for compliance and audit purposes, it's important to keep track of any changes a person makes to a record.

AI and human refinement, along with these principles, are table stakes when it comes to selecting an AI-native MDM solution that will not only solve the hard problem of entity resolution, but also turn your data into gold…golden records, that is!

Tamr's unique identifiers solve the challenge of data fragmentation. **Watch this video** to see how the TamrID works.

8

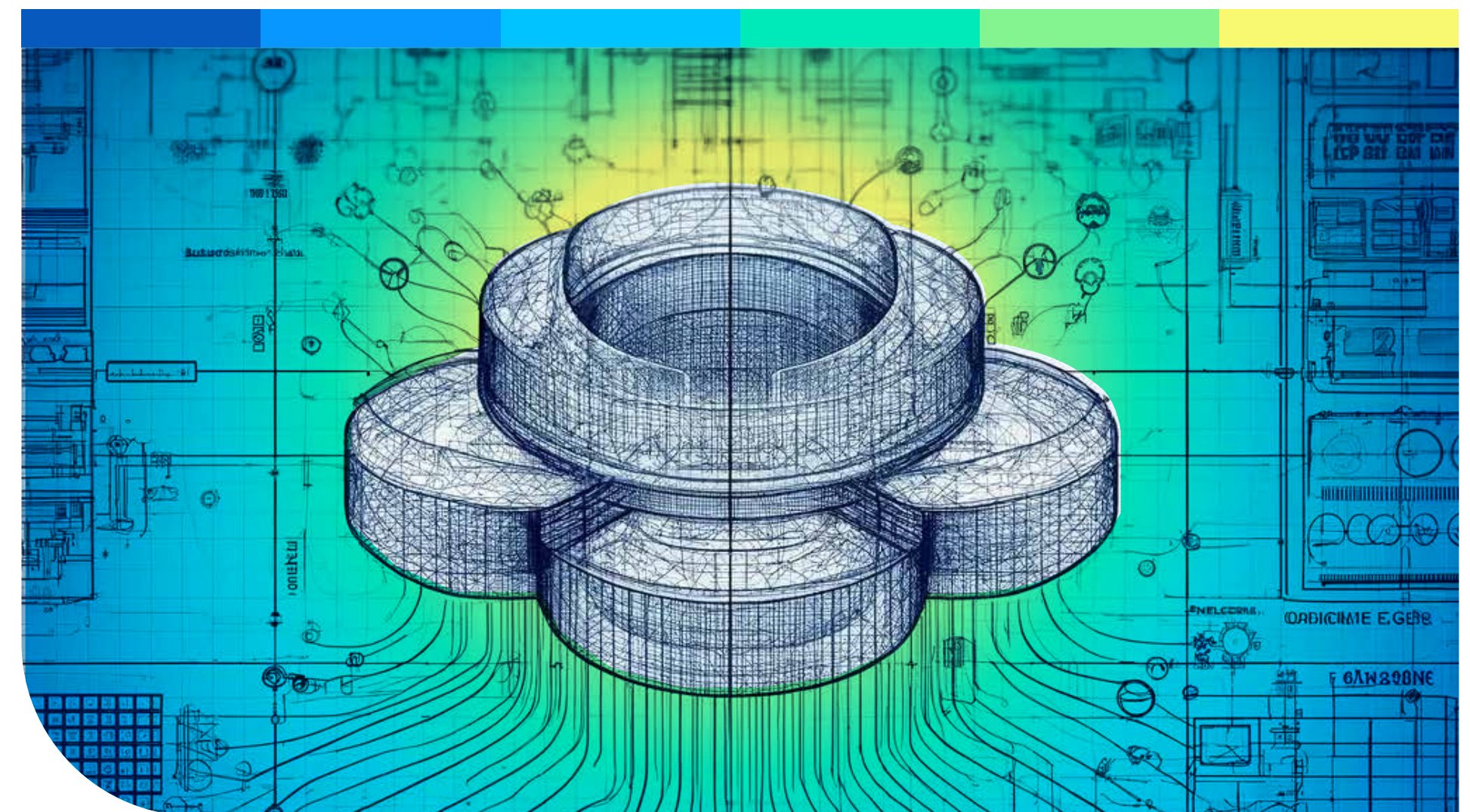**Match verification with external data:** incorporating external reference data for matching isn't just a 'nice to have' if you want to maximize match rates and accuracy; it's the only way to verify that an entity exists in the real world (essential for use cases like KYC and fraud detection) and gives you the best chance of finding 'hidden' connections in your data, such as those from a company name change.

9

Enrichment: the best version of your data likely lives outside your firewall. Solutions with data enrichment capabilities integrate your internal data assets with external data to increase the value of these assets. And, they add additional relevant or missing information so that your data is more complete and usable.

10

Real-time: duplicate data is a pervasive problem. That's why you need to stop it in real time. Using real-time capabilities, the best solutions search the database before creating a new record. If a matching record exists, they update it. If not, they proceed with creating a new one.

Golden Records: The Future. Powered by Tamr

Tamr uses AI to eliminate multiple versions of the truth. **Watch this video** where Tamr's Head of Strategic Solutions, **Ravi Hulasi**, shows you how Tamr delivers accurate, up-to-date views of business entities.

If you want to deliver the golden records your business needs to make better decisions, Tamr is here to help. By employing AI to solve the problem of entity resolution, Tamr's unique AI-native approach is built to help businesses like yours finally realize the promise of golden records.

Unlike rudimentary tools of the past, Tamr uses advanced AI to automatically identify and resolve inconsistencies across data sources. With Tamr, businesses can finally deliver the golden records every stakeholder needs to realize value from their data.

Data engineers

When it comes to matching and resolving entities across systems and sources, anomalies and data quality issues cause challenges for data engineers. Inadequate tools, such as legacy, rules-based MDM solutions, make it difficult to reference other sources and validate attributes, leaving data engineers lacking confidence in their ability to match entities and create golden records.

In contrast, Tamr infuses AI into data pipelines, making processes such as entity resolution, feature extraction, and data labeling more efficient and effective. Persistent Tamr IDs connect entities across sources so it's easy to track lineage and provenance, providing insight into the data's history as it changes.

Further, Tamr provides verified matches to external, referential sources so data engineers can feel confident that a company, supplier, or contact exists, and that alternate values such as different names or email addresses are valid.
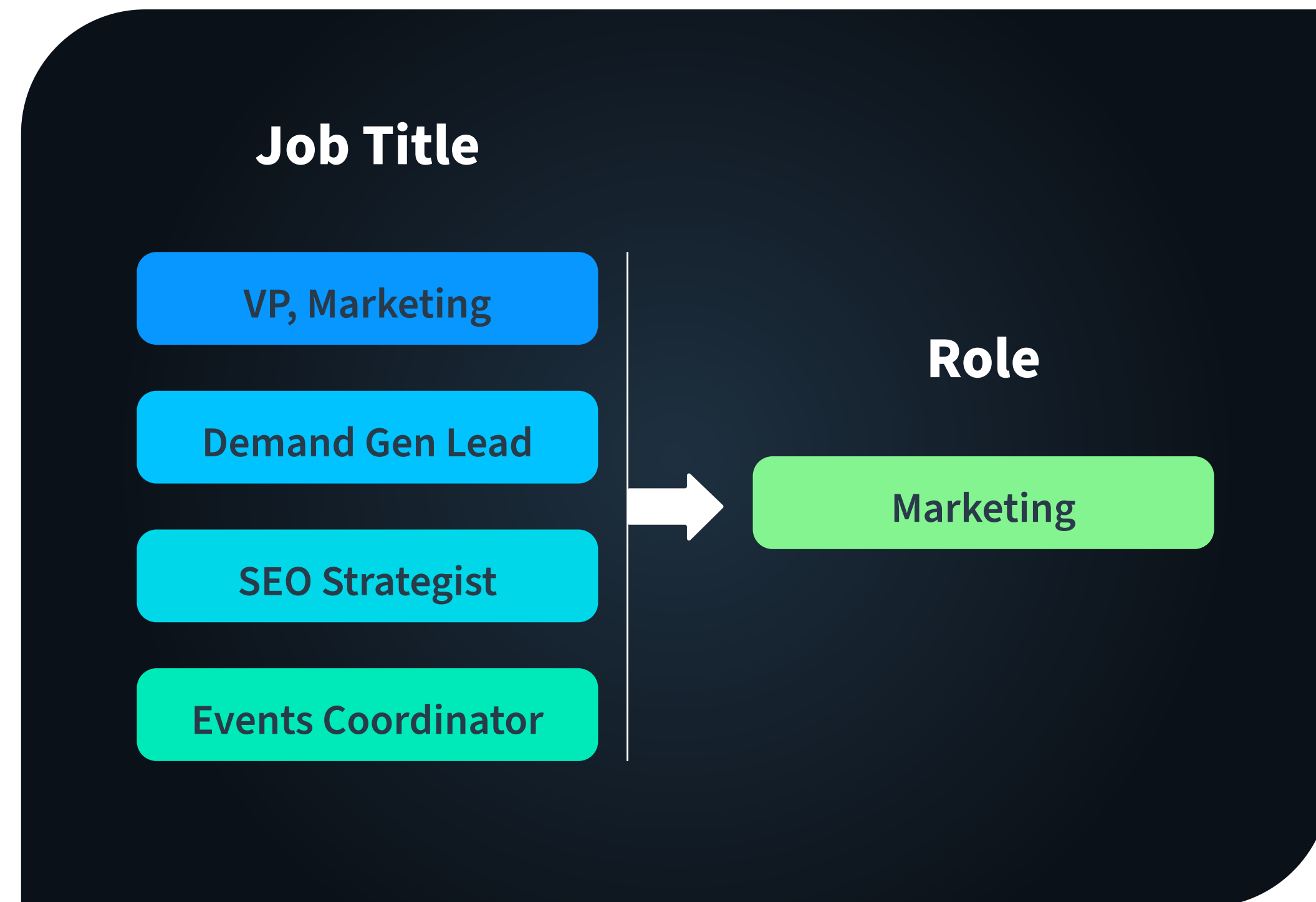

Clean   **Verify**   **Match**   **Enrich**

Tamr also allows data engineers to classify individual values into broader categories, increasing the quality and completeness of their golden records. A great example is grouping similar job titles into a broader category defined by a role.

As these attributes change, or new data comes onboard, AI actively intervenes to continuously test and standardize raw inputs using a combination of reference data, pre-configured business logic, and large language models (LLMs). As a result, AI can match the new data to existing entities in **real time**, preserving and enriching the integrity of the golden records.

Job Title

VP, Marketing

Demand Gen Lead

SEO Strategist

Events Coordinator

Role

Marketing

Crack the code on entity resolution with Tamr's Chief Architect Emeritus **Nik Bates-Haus**. **Watch this five-part video series** for an in-depth look at how Tamr's innovative technology solves entity resolution at scale.

Data consumers

Data consumers include the data analysts and business end users who turn valuable information into the insights, knowledge, and decisions that move the organization forward. Finding the data they need – and trusting the results of their search – are the biggest obstacles data consumers face. Many times, they hunt for data but can't find it because the search string they are using doesn't recognize the data they need. And even when their search returns a result, they remain skeptical, wondering if the insight they've uncovered is valid, correct, and complete.

With Tamr, data consumers benefit from AI-powered search which serves up data on their terms. Multiple views of the customer highlight what the data consumer cares about - and only what they can access. They can view the history of the data and not only see what changed, but also how it changed, providing the critical context

they need for informed decision-making. And they can provide feedback, alerting data stewards when they suspect that results are faulty.

When data consumers successfully find the data they need, they become advocates, touting the benefits of data and its positive impact on the business. As a result, culture begins to shift towards one where data is honored and its value proclaimed as the reason why the business is thriving.

Data stewards

If you're looking for a data steward, odds are you'll find them buried under a mound of data. Keeping pace with the never-ending barrage of data is exhausting, and with a limited number of data steward resources, they can only review so many records at once.

Preserving data integrity is a data steward's job.

But when data is constantly changing, curating data and improving its quality becomes a full-time effort, leaving little time to add value through human refinement and feedback.

That's where AI comes in. Tamr does the heavy lifting when it comes to surfacing potential matches or possible anomalies within golden records. It spots records that are potentially mismatched and inconsistencies obscured by dirty data so that data stewards can resolve the issues and deliver high-quality, trustworthy golden records. Using AI, data stewards become more efficient, freeing up time to focus on more strategic tasks.

Furthermore, when AI surfaces anomalies, data stewards can learn from the AI and gain a better understanding of what an anomaly looks like so they can more easily spot them as they fill in missing data attributes.

In addition to its innovative technology, Tamr also factors in an invaluable element AI simply can't replace: **human refinement**.

Human refinement validates golden records by providing context and insight AI can't deliver, ensuring that the golden records are the highest quality, most trustworthy version of key business entities.

Human refinement is an indispensable tool for reliable, golden records. To discover why humans are so critical to developing golden records, read **Power to the People: The Role of Humans in AI.**

Furthermore, Tamr offers speed, scalability, accuracy, and durability that are **beyond compare**. Traditional MDM solutions promised they would deliver "golden records," but in reality, they fell short. Their rules-based approach, built for static data, can't keep up with the pace at which modern data is growing and evolving. Writing, modifying, and maintaining the rules is tedious and time-consuming, requiring an inordinate amount of human effort that simply isn't scalable.

Tamr's AI-native MDM overcomes the limitations of legacy, rules-based MDM solutions by giving modern, data-driven businesses the flexibility to adapt as their needs and strategies change. It puts the management and control of data into the hands of people who need it: the data engineers, data consumers, including data analysts and business end users, and data stewards who use data to drive business growth.
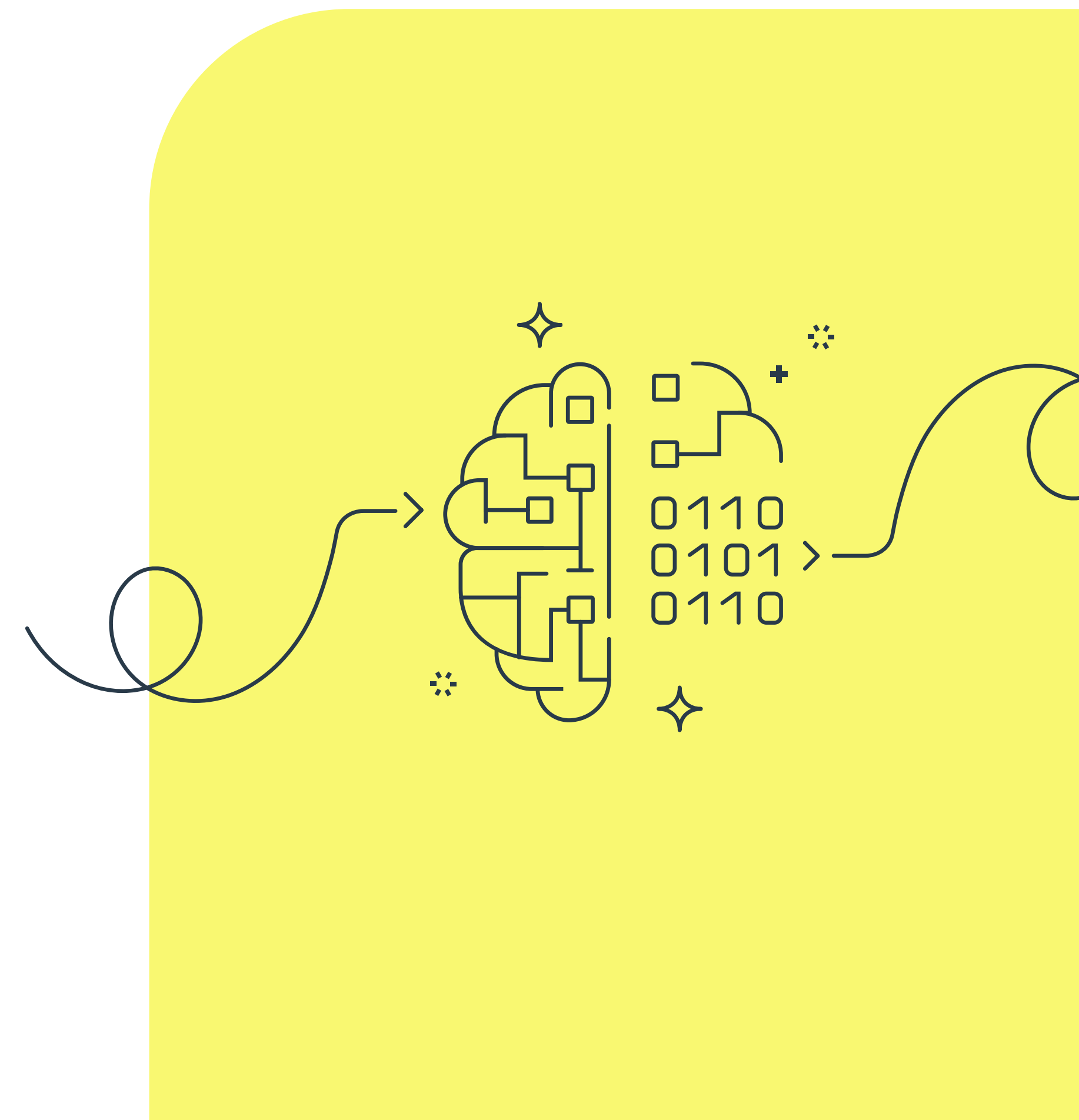
The Golden Records Effect: How Tamr Helps Customers Drive Business Impact Fast

Overcoming the challenges associated with entity resolution isn't easy, and it's rare to find organizations that have successfully done so. Until now.

Let's explore how three Tamr customers delivered meaningful business impact quickly and at scale using AI-native MDM to deliver golden records.

Western Union

▶ Customer Success Video

See how they did it.

Read the full case study.

TOYOTA

▶ Customer Success Video

See how they did it.

Read the full case study.

OLDMUTUAL

▶ Customer Success Video

See how they did it.

Read the full case study.

A Golden (Record) Opportunity

As proven by Western Union, Toyota Motors Europe, and Old Mutual, once companies realize the promise of the elusive golden records, everything changes. Powered by AI and refined by humans, AI-native MDM makes it possible to finally deliver the golden records companies need. By combining embedded similarity with human feedback, AI delivers best-in-class match rates with external data so that users can feel confident their data is accurate and reliable.

Golden records are tailored to the consumer as well, integrating every identifier in the system with human validation to create a personalized, single view of every customer. And because AI-native MDM solutions learn and improve from machine-generated feedback, data management becomes more efficient and adaptable over time, ensuring that the golden records will evolve as business needs change.

Golden records deliver better, more comprehensive views of key business entities, creating positive ripple effects for businesses striving for data-driven collaborations. Hidden insights, previously obscured by dirty data, reveal themselves, prompting everyone to make informed decisions with greater speed and precision. And because users have a single, authoritative, accurate version of key business entities, they can better identify emerging trends and personalize customer experiences in ways that align with today's customer expectations.

Golden records foster interdepartmental collaboration by aligning teams on a unified version of the truth, ensuring consistent management of customer experiences across the various touchpoints, and improving analytics and decision-making to drive customer loyalty.

The Future of Data:
Golden Records as the Foundation for Enterprise Success

The future of data management promises to be dynamic and transformative. Emerging trends, technologies, and innovations like AI are reshaping how organizations use data to drive their businesses forward.

As organizations continue to grapple with the size and complexity of modern data, AI will secure its place as a critical tool in the modern data ecosystem. The reality is that big data is getting bigger. And with it, the proliferation of data sources is expanding exponentially. The only feasible way to keep up is by embracing AI and using it to create golden records that extract valuable insight, enable greater flexibility and scalability, and drive true data-driven decision-making.

With Tamr's AI-native MDM at the forefront of data innovation, organizations worldwide will reap the benefits of golden records and the holistic view of key business entities they provide. By using Tamr to elevate their customer experiences, drive operational efficiencies, and uncover new opportunities to grow, companies worldwide are changing the data management game and positioning their business to grow and adapt today and into the future.

Discover the value Tamr can provide for you. **Book a demo** to learn more.

Tamr

Tamr provides the only AI-native Master Data Management (MDM) solution that delivers real-time master data for every dashboard, application, and person in your business. Tamr accelerates the discovery, enrichment, and maintenance of golden records, enabling informed decision-making, improved revenue growth, and better customer experiences.

Tamr's patented, AI-centric approach – with human refinement and oversight – delivers value in days or weeks, not months or years like traditional rules-based MDM and DIY solutions. And with intuitive Customer 360 pages, your business can improve data accessibility across the organization and leverage the best, most accurate data to support analytical and operational use cases in real time.

Learn more at **tamr.com**