

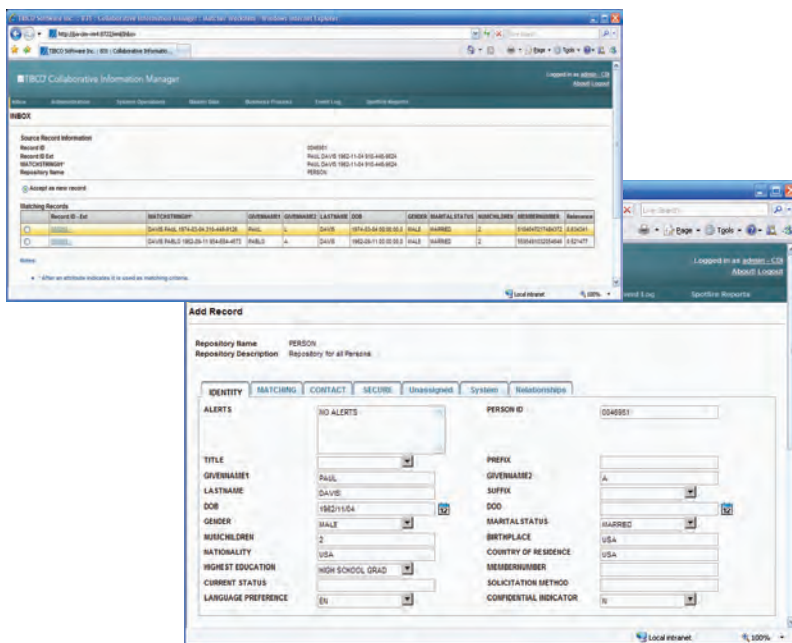
Managing Data Quality with TIBCO Collaborative Information Manager

Data quality (DQ) refers to technologies used for validating, cleansing, and de-duplicating data. For example, customer data in CRM and other systems is often inaccurate, incomplete, and inconsistent. This leads to errors in billing, customer service, and marketing. In addition, the data is full of duplication, leading to unnecessary spending on marketing campaigns and fulfillment. Poor data quality makes it difficult for organizations to calculate customer profitability and assess risk exposure.

Master data management (MDM) manages information about products, customers, vendors, and other core data assets. An MDM system ensures that master data is accurate, complete, and consistent. To achieve this, MDM solutions must manage the information lifecycle. Cleansing, validating, and de-duplicating that information is a core component of that lifecycle.

TIBCO Collaborative Information Manager™ software solves the problem of de-duplication through advanced match-merge capabilities.

Matching two pieces of information to identify potential duplicates is managed through an embedded match engine. TIBCO's match engine, based on the Netrics Match Engine, uses advanced mathematical algorithms to mimic human similarity assessment – essentially making decisions the way a human would make them. This revolutionary approach eliminates the need to model and maintain scores based on elaborate rules that try to do the same thing.



TIBCO Collaborative Information Manager software provides a rich data stewardship GUI for match selection and merging of potentially duplicate records.

AT A GLANCE

TIBCO Collaborative Information Manager software, TIBCO's multi-domain MDM application, delivers embedded match-merge technology to identify and eliminate duplicates in all master data. TIBCO embeds the Netrics Match Engine, which uses advanced mathematical algorithms that mimic human similarity assessment rather than the traditional rules-based approach.

BENEFITS

- **Accuracy:** Advanced match technology produces very accurate results, leveraging over 20 years of research in the field of inexact-matching.
- **Faster to Implement:** The no-rules approach requires less services effort, is faster to implement, and has a lower total cost of ownership as it does not require extensive re-work for additional data types or systems.
- **Configurable:** The match-merge activities are configurable, so users are able to determine match thresholds, merge precedence, and other parameters necessary to ensure accurate, duplicate-free information.
- **Scale:** The computational efficiency and in-memory architecture yields a very highly scalable engine.
- **Multi-language:** The match engine is multi-language, a key requirement for organizations that have customers around the world and must perform matches on names and other fields in a variety of European and Asian languages.



Managing Data Quality with TIBCO Collaborative Information Manager

ATTRIBUTES & CAPABILITIES

No-Rules Approach to Matching

In the traditional rules-based approach to matching, IT works with business analysts to codify, in rules, how an analyst would determine a match. This can be a lengthy process. Because no two data sets are identical, these rules must be adjusted per data type, and often per source system. Leapfrogging that approach, TIBCO's match engine, based on the Netrics Match Engine, uses advanced algorithms that eliminate the need for creating, tuning, and maintaining rules. Using techniques developed through 20 years of theoretical and systems development, the TIBCO match engine determines the probability of a match by applying a human-style assessment of similarity. At the core are algorithms that produce a numerical measure of similarity between data, based on an optimal weighted bipartite matching of letters and polygraphs. This fundamentally different approach captures a more flexible and more "human" notion of similarity than alternative approaches. The result is an extremely intelligent search.

Multi-Domain

One of the biggest shortfalls of the rules approach to matching is the rules must change with the type of data and system. The TIBCO approach to matching is multi-domain, requiring little to no tuning regardless of data type or language.

Rules-Based Workflow

Matching and merging are implemented as activities within the context of TIBCO's rules-based workflow architecture. The configurable workflow allows customers to configure how matching should be performed, balancing performance and accuracy by applying multiple match patterns of varying stringency, and applying configurable rules to make decisions based on the result. Matching and merging are thus not isolated activities but built into information lifecycle processes.

Merge and Un-Merge

Once two records have been identified as a match, they must be intelligently merged and linked in order to create a single golden-copy record, while often preserving the source information. TIBCO provides a precedence-based approach to merging. The application keeps track of which source system has priority for which attribute and then applies that logic to determine survivorship. IDs of the source system can be passed into the golden-copy record to maintain links between the golden-copy and associated source systems. A full data lineage trail is maintained and can be queried from the UI or reported against. This lineage information is used to un-merge the data. A simple un-merge can be performed from the GUI, while more complicated un-merges can be rules or workflow-based.

Data Stewardship

Even the most efficient algorithm will not be able to automatically determine every match. That's why TIBCO augments the match engine with rich data stewardship capabilities. Data stewards can, through an out-of-the-box GUI, make decisions on matching and merging information that fall outside the engine's configurable match thresholds.

High Error Tolerance

Manually entered data is typically full of errors such as typos, transpositions, out of order words or fields, missing spaces, etc. TIBCO matching technology can account for such errors in both the query and match record and still produce very high quality results, eliminating the need for extensive data profiling or pre-processing.

Highly Scalable

Matching is performed in memory, and the engine is highly computationally efficient, allowing for sub-second response times on very large data sets.

TIBCO Software Inc. (NASDAQ: TIBX) technology digitized Wall Street in the '80s with its event-driven "Information Bus" software, which helped make real-time business a strategic differentiator in the '90s. Today, TIBCO's infrastructure software gives customers the ability to constantly innovate by connecting applications and data in a service-oriented architecture, streamlining activities through business process management, and giving people the information and intelligence tools they need to make faster and smarter decisions, what we call The Power of Now®. TIBCO serves more than 3,000 customers around the world with offices in more than 20 countries and an ecosystem of over 200 partners. Learn more at www.tibco.com.

©2009, TIBCO Software Inc. All rights reserved. TIBCO, the TIBCO logo, The Power of Now, TIBCO Software, and TIBCO Collaborative Information Manager are trademarks or registered trademarks of TIBCO Software Inc. in the United States and/or other countries. All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only. 09-10333



www.tibco.com

Global Headquarters
3303 Hillview Avenue
Palo Alto, CA 94304

Tel: +1 650-846-1000
+1 800-420-8450

Fax: +1 650-846-1005