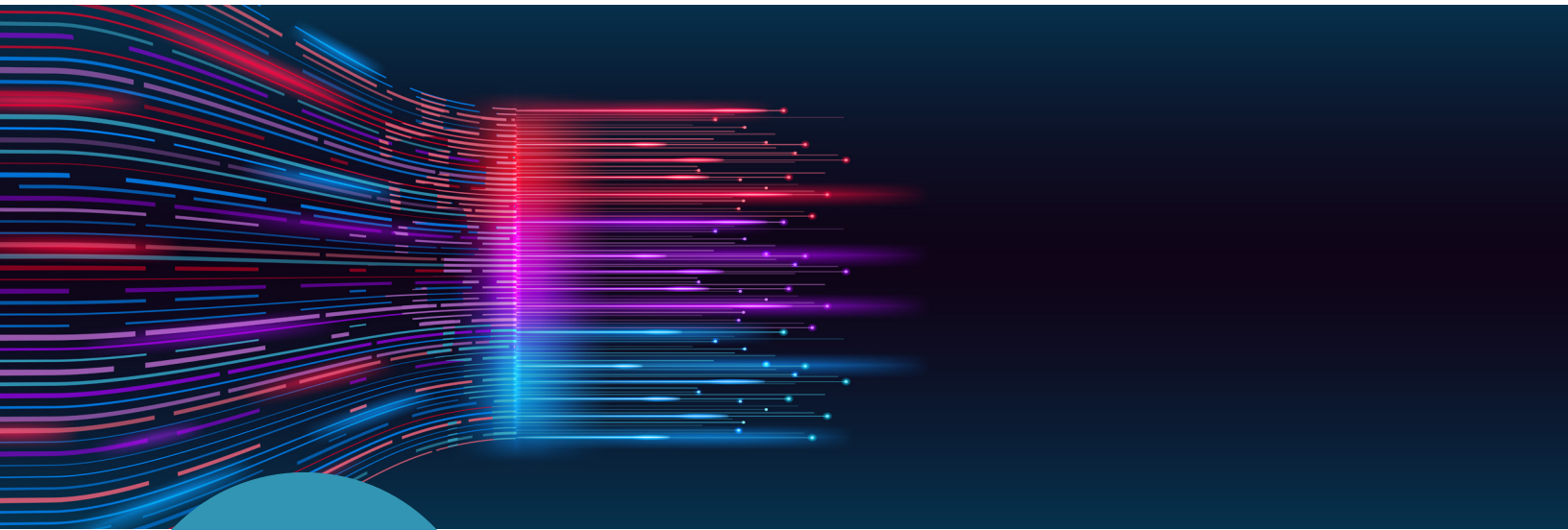# The Pathology of Matching

BY JOSEPH SANTOS
DUN & BRADSTREET

> "Everything is pathology,
> except for indifference."
>
> – EMILE M. CIORAN

The practice of pathology studies the causes and effects of disease in the medical sciences. I've always thought that the same concepts could be applied in business to the understanding of our data. For example, what are the underlying reasons that data is or isn't helping sales grow? Studying our data, whether or not we've identified a problem, can help us understand how it is, isn't or could be contributing to our organization's goals. Applying Cioran's quote, we should always be studying and diagnosing our data; otherwise, we could be accused of not caring about this valuable asset.

Data stewards and guardians must constantly study all aspects of their data, being on the lookout for patterns, similarities, and deltas to continuously derive value. Studying the organization, processes, and usage of our data is integral to ensuring that it is working favorably for our business, across teams, departments, and systems. In most cases, this begins with matching.

MATCHING IS THE FIRST STEP

As data practitioners, we rely on outside data to enrich our own data assets and keep it up to date, complete, and relevant. To accomplish this, we go through a process that typically includes working with a trusted outside data provider (source data), and aligning that external data with our internal datasets through a series of attributes. This exercise is known as matching, and its goal is to match our

internal record with the same (or similar) entity in the external, more robust dataset, thus acquiring the pertinent attributes for data enrichment and augmentation. We might even find opportunities to further refine our internal data through this exercise, e.g., discovering duplicates or inaccurate entries. When a match is executed and we are able to source another entity, our processes operate as designed and life goes on. But what happens when some records do not match?

Not every record is going to match, and there are many reasons for that. On the client side, data readiness and completeness are the usual culprits. For example, if the "city" field is one of the attributes needed to match, but it is not fully populated in the client dataset, our opportunity to match and enrich those records will decrease. On the vendor side, for example, the size of their referential data – i.e., how deep is their data well? – may impact match performance. Regardless of the reasons behind match performance, it's important to understand that a "no-match" record is as important as a matched record. This is because failed attempts carry significant forensic intelligence. Once we've ruled out systemic and delivery challenges, we can dig into the main reasons for no-matches.

No-match records hold opportunities to elevate discussion that may lead to new solutions or even new discoveries. It's an opportunity to revisit current consumption of external data as well as internal business processes. It may even be an opportunity to review vendors' offerings.

---

1. For an excellent discussion on this topic, see "Straight Talk on Matching; Why 100% Resolution is Unrealistic and May Be Counterproductive."

CREATING OPPORTUNITY IN NO-MATCH

When no-matches become an issue with your data augmentation and enrichment initiative, what can you do to learn from it? As I see it, we have three opportunities to turn these events into actionable solutions.

1.	2.	3.
DATA READINESS	**MATCHING IS A TACTIC. IDENTITY RESOLUTION IS THE STRATEGY**	**NO-MATCH RECORDS CAN UNCOVER INTERNAL BUSINESS TRANSFORMATIONS**

1. DATA READINESS

Your data has an important role to play in your business goals, so understanding the process by which data is managed can bring further efficiencies to your data journey. Data readiness is the mindset embedded in the business where everyone understands and respects the value of using data to help navigate the business and open doors to opportunities. No-match result analysis may reveal contexts that pertain to your data readiness.

Matching is always going to be a two-way street. The external vendor/partner reference data you select should be robust enough to positively impact your matching goals and structured enough that it meets your purpose. Dun & Bradstreet's unique Live Business Identity can be considered a living, breathing descriptor of almost every business on Earth, providing a comprehensive and continuously updated view of your business relationships. Use this to your advantage.

You can use no-match results to help identify areas for improvement within your dataset that impact master data management.

COMPLETENESS – Let's say you are using address components (street, city, state, etc.) as matching attributes. For your no-match population, understanding the number of records that are missing these field values can surface an area of improvement in your data gathering process. Perhaps in the past those fields were merely a "nice to know," but now with account based marketing and go-to-market changes, your business might rely on those values. Our recommendation would be that these fields should be upgraded to become required at time of creation and implement data management support needed to enforce this new requirement.

QUALITY – In the event there are values populating crucial fields/attributes, scrutinizing them in the no-match results will reveal patterns and data present through a rudimentary analysis of the values themselves. You could find incorrect values that may have been created due to a departure from protocol or data that should not be in there, such as postal code values in the city field or telephone numbers in "Address Line 1." Your findings can serve as notification that your data capture processes need attention and improvement.

AGE – On the surface, the age of your data might not be apparent, but it can produce no-match results. This is not uncommon. If the dataset you're matching to includes only current and active businesses, data for companies in your database that are out of business will fall under a no-match scenario. For business records that appear to be legitimate, create a random sample and research their eligibility. Once you find that these are older and unneeded legacy records, you can have an internal conversation about archiving. Some Dun & Bradstreet clients ask for a second pass matching with the out-of-business dataset. Apart from adding further intelligence on the status of your organization data, this type of strategy can go beyond data quality tactics and extend to fraud prevention by interrogating transactions and their timing when related/linked to these out-of-business records.

Through a preliminary assessment of your data and analysis of the findings, you or your data partner(s) should be able to provide a cursory understanding of your data readiness. Arguably, the no-match results could also provide you a decisive current state scenario of your data readiness.

2. MATCHING IS A TACTIC. IDENTITY RESOLUTION IS THE STRATEGY.

Matching is a tactic; a tool. It has its limitations if used solely as the strategy. Finding the truth goes beyond a single customer record. Expand this methodology; go beyond your traditional enrichment and augmentation process. Traditional matching is likened to having facial recognition on your mobile device that unlocks when focused on only one acceptable angle or expression of your face. You will have limited success. The face can display multiple expressions and be shown in different lighting and angles. The strategy of using facial recognition for security lies in its effectiveness to identify you (and only you) as the user but through your many dimensions. Your master data is no different. Identity (or entity) resolution is the process whereby multiple sources, datasets, and versions are used to resolve the identity of targeted entities. Opportunities for identity resolution can manifest in no-matches as indicated by these examples:

BUSINESS NAMES – These are likely to have many iterations such as the legal business name, local language, a D/B/A, trade style, etc. There will also be many permutations on how this attribute can manifest. These may live in many different record sets throughout the enterprise data ecosystem.

ADDRESS OR LOCATION – Address data also comes in many forms. A singular business entity could have a legal address, a secondary address, a mailing address, a vanity address, etc. It also could have an alternative address if the entity is large enough to occupy multiple streets of the block.

PAST VERSIONS – Entities may change locations, change names, face divestitures (mergers and acquisitions), or go out of business. Master data is alive, so to speak. Which version do you have? Perhaps what you have is a combination of the above.

DIGITAL PRESENCE – For starters, we're talking about URL domains and IP addresses, and just like the other examples, there can be multiple versions of these. These digital attributes can be found as visitor data at your digital front door. It's another "facial angle" or "facial expression" to consider.

KEY BENEFITS OF DUN & BRADSTREET'S IDENTITY RESOLUTION METHOD

- Assess your current matching process
- Identify and benchmark current data quality issues
- Grade your B2B match reliability

- Illustrate process improvements by applying a quality referential database standard
- Improve account relationships by establishing a quality data foundation built on a proven process

DETERMINES UNIQUENESS OF ANY GIVEN ENTITY BEFORE ASSIGNING D-U-N-S NUMBER

IDENTITY RESOLUTION YIELDS SINGLE VIEW OF THE ENTITY

SOURCE 3
A.B.C.
123 Elm St.
Bethlehem, PA 18025
(SIC) 5411-02
Chuck Smith, President
D: OX20PH
$3,000 credit limit
$1,100 outstanding and 12 days past due

SOURCE 2
Alpha Plastics
P.O. Box 111
Bethlehem, PA 18055
Chuck Smith, President
610 882-7600
ID: A595660662
Product code A on 7/2013
Product code B on 9/2011

SOURCE 1
Charles Smith
123 Elm St.
Bethlehem, PA 18025
(SIC) 5541-99
215 882-7600
ID: 001200000000wXgAAI
SMB channel customer
Last touch point on 6/2013

D&B D-U-N-S NUMBER: 12-345-6789	
Legal Name	ABC, Inc.
Tradestyle	Alpha Plastics
Address Postal	P.O. Box 111 Bethlehem, PA 18055
Address Physical	123 Elm St. Bethlehem, PA 18025
Phone	610 882-7600
SIC	5411-0202, 5541-9901
Primary Contact	Charles Smith, President

CONSOLIDATED PROFILE
D-U-N-S assigned
Names rationalized
Addresses categorized

Identity resolution is best attained when going beyond the traditional single-record-pass match. Consider the above examples. While facial recognition software will obviously not work for business entities, Dun & Bradstreet's patented identity resolution method, using Live Business Identity, functions the same way. Its referential data is inclusive of all the examples listed above and more. Through complex engines and iterative processes, it reconciles your data with the robust catalog of its many dimensions. The idea is to use the many facets and stages of your data to provide the intelligence you need.

The opportunity here is to establish identity resolution as a unified process. The process of enhancing the intelligence of your internal master data should not be a siloed exercise. Companies that are tactical in approach might have their operations set that way, and are thus not maximizing their efficiency. Simply stated, identity resolution is an end-to-end process. It should be inclusive of web, automated, and manual inbound data; master data management; lead management; analytics operations; and sales and marketing outreach.

Just as personal identities can change over time (for example, by marriage or other name changes), data is dynamic as well. Though identity resolution uses matching fundamentals, it is a multidimensional process. It increases the effectivity of your data enrichment initiatives far beyond relying on the properties offered by a single-record-pass match.

> "Identity resolution is an end-to-end process. It should be inclusive of web, automated, and manual inbound data; master data management; lead management; analytics operations; and sales and marketing outreach."

3. NO-MATCH RECORDS CAN UNCOVER INTERNAL BUSINESS TRANSFORMATIONS

This opportunity is a bit more complex and requires a specific example. A Dun & Bradstreet customer, a leading computer hardware company, noticed their match rates dropped significantly to what seemed an alarming level. There were no big changes noted by the data team for both data and processes. A large portion of the no-match population had significant revenue attributed to them. The addresses were valid, and the business names were recognizable as large companies. These records should have found matches within the D&B reference data. An analysis of the no-match population showed that though the business names and addresses were valid, those businesses did not exist in the given address locations. It turned out that the addresses belonged to known data center institutions that catered to those businesses. The analysis revealed that our client had shifted to a new channel-selling method which affected the data capture. The net result was a shift in the data capture and data management that allowed our client to identify these co-location transactions for data management and analytics purposes.

YOUR DATA NEVER STOPS EVOLVING

At its very core, managing data is really about managing processes. The data is merely the manifestation of the process and rules that exist to generate and maintain it. No-match results can provide these revelations.

No-match forensic analysis is key to understanding your data. The success of the analysis lies deep in the data and could be uncovered by the match/no-match results with your data context and a strong reference source such as Dun & Bradstreet.

Just remember, it takes two datasets to dance the matching tango. The variability of customer data can help match conversations and potentially business processes.

The last and most important opportunity of all: Data is alive and fluid; it can change results without you, the data practitioner ever knowing. Know that data is always changing. The pathology of no-match data will provide you the insights you need to act.

> "Managing data is really about managing processes. The data is merely the manifestation of the process and rules that exist to generate and maintain it."

About the Author

A 23-year veteran in the data and technology space, Joseph Santos is a Principal Consultant on the Dun & Bradstreet Data Advisory Services team providing strategic guidance on data initiatives to organizations, large and small. Joe was also a long-time Dun & Bradstreet customer working in the trenches as a data leader for several high-tech companies. His extensive experience and expertise allow him to deliver strategic solutions on Master Data, hierarchies, customer data management, data governance, data stewardship and data forensics.

About Dun & Bradstreet

Dun & Bradstreet, a leading global provider of business decisioning data and analytics, enables companies around the world to improve their business performance. Dun & Bradstreet's Data Cloud fuels solutions and delivers insights that empower customers to accelerate revenue, lower cost, mitigate risk, and transform their businesses. Since 1841, companies of every size have relied on Dun & Bradstreet to help them manage risk and reveal opportunity.

dnb.com | Twitter: @DunBradstreet