Data Ladder
Get the most out of your data

White Paper

Creating the Golden Record with Data Ladder's Merge Purge Feature

68 Bridge, St. Suite 307
Suffield, CT 06708

+1 888-779-6578

Sales@DataLadder.com

www.DataLadder.com

As much as companies enjoy basking in big data, they still lack the capacity to manage, store and handle data. For most part, data is stored in silos. Teams and departments work in isolation. A central data management framework is missing. Companies do not have a complete, consolidated view of their customers.

The result of such disparate data storage is the lack of a unified source of truth that disrupts real-time engagement and personalization goals. Many organizations recognize the need for personalization, but fail to execute the plan, for the lack of coherent data. It's virtually impossible to deliver value when customer data is dumped in data lakes or scattered across multiple systems, accessed and operated by multiple users.

To succeed with any kind of customer personalization or satisfaction goals, firms need access to unified, consolidated records. The term for this kind of consolidated record is, 'golden record.' This whitepaper will guide you towards making golden records a possibility. It's a high ambition, but one that can be done if you have the right team, the right tools and the right process.

But before jumping into the how, it's important to understand the challenges firms are facing. One of the most critical is data quality. Failure to ensure data quality can render all data, whether big or otherwise, virtually useless because of inaccuracies and the fundamental unreliability of the insights they are bound to yield. In this regard, data quality is a vital prerequisite for any sort of analytical insight or customer personalization goals.

The State of Data Quality in Companies Today

Until recently, companies have been investing millions of dollars in acquiring, 'more' data, but more data is hardly helpful if it cannot be processed and used for its intended purpose.

With the state of data quality being as it is, companies are finding it increasingly hard to make sense of their data. To engage the audience with relevant offers and recommendations, companies need accurate, complete, reliable customer data. While data hygiene remains a critical challenge (as it always has been), it's the consolidation of data from disparate sources that is the bigger concern.

Why the Need for Consolidation?

On average, an enterprise is connected to 461 apps and systems. Every day, these systems generate gigabyes of data, varying in type, purpose and format. There will be data from social media applications, mobile logins, ad campaigns, and other application functionalities. It is imperative for the organization to clean, merge, purge and consolidate this data to achieve business goals.

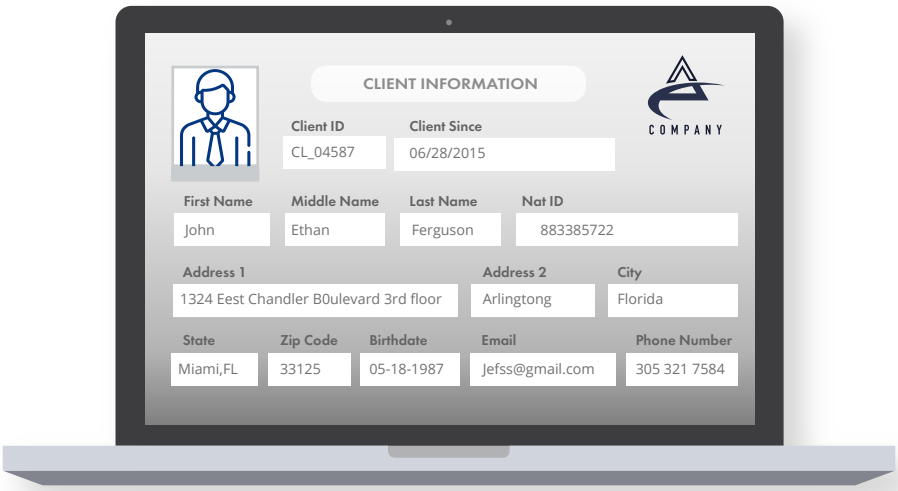
For instance, if the goal is to launch a new loyalty card program, the company must decide on the audience segmentation (customers who've been with the company for five + years), the information for that audience (transactional data) and their social media accounts (for targeted advertising). To carry out this activity, the company will need to have a consolidated version of their customer data. If they don't already have clean, consolidated data, they'll end up spending months in getting the data ready for this goal.

Time in today's fluid and dynamic market is limited. If 80% of a business team's time is spent on just preparing data, companies will find it hard to meet their goals.

Data consolidation will enable firms to improve its data quality, get accurate data for business reports and analytics, understand the customer journey, improve sales performance, improve customer satisfaction and much more. Creating the golden record will pave the way to taking a competitive lead.

What is the Golden Record and How Do You Overcome Challenges

The Golden Record is defined as a single, well-defined version of all the data entities in an organizational ecosystem. Essentially a Master Data Management (MDM) concept, it has now become a vital component of business processes, specifically of marketing and sales.



In-Store

Customer ID	CL_04587
Name	John Ethan F
National ID	883385722
Address 1	1324 Eest Chandler Bvd
Address 2	Arlington¶
Email	jeffss@gmail.com
Birthdate	05/18/i987

E-Commerce

First Name	Jhon E.
Last Name	Ferguson
Address	East Chand. Blvd 3rd Floor
Zip Code	33125
State	Miami
City	Florida
Client Since	06/28/2015

Customer Support

First Name	Johny
Last Name	Ferguson
Address	Arlington Blvd 3rd Floor
Zip Code	3053217584
State	Jeffss#gmil.com
City	NULL
National ID	883-385-722

Also known as the, 'single version of truth,' a golden record is the most accurate, complete, updated information of an entity. For modern businesses aiming to up their marketing or sales strategies, and adopting customer-centric approaches, this golden record is of immense value.

Despite being useful for businesses, the Golden Record is an immensely difficult task to undertake. You will need to deal with underlying data quality challenges before you can even consolidate data good enough to make the record.

We like to call these challenges as the, 'Three Bad Ds of Data,' that need to be resolved before pursuing further goals.

1. Dealing with the Duplication Plague: Take a look at the image below.

That's two records of a single entity, with one being a CRM record where the phone and address data has changed. The other is a more recent one from the ERP where there are a new phone and address data. If you were to analyze this data, how would you know which of it is the correct one? This is just one example. What will you do if this duplication becomes exponentially worse, with every new duplication requiring a verification and validation process that consumes valuable resources?

The first step to data consolidation lies in resolving the duplicate problem. There is no room for compromise when it comes to duplicates. They stay in your database for years, eating up valuable space and causing havoc every time you have to run a report or use the data for a business purpose.

CRM

First Name	Jonathan E.
Last Name	Ferguson
Address	East Chand. Blvd 3rd Floor
Zip Code	33125
Phone	516 468-7274
City	Florida
Client Since	06/28/2015

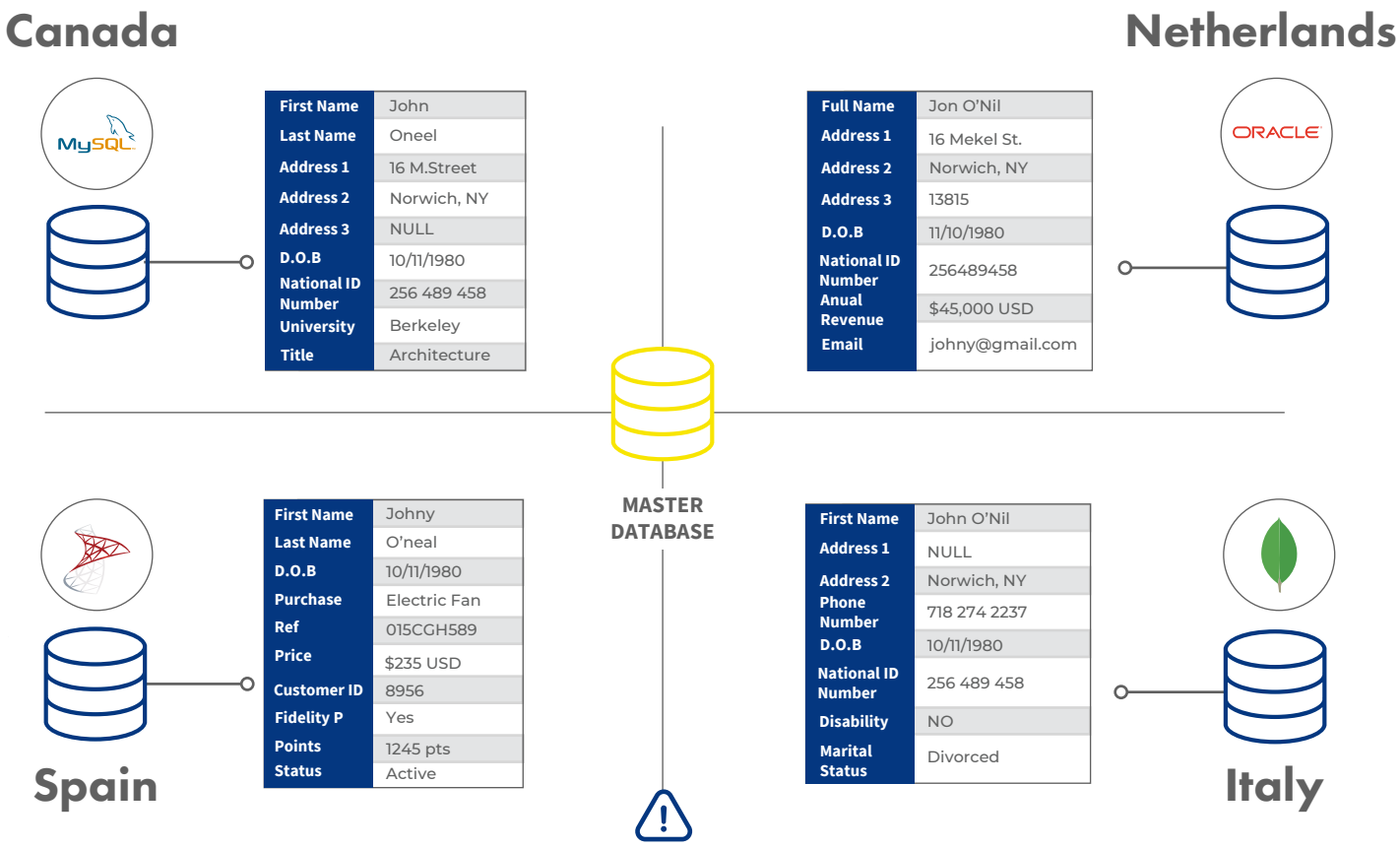
ERP

First Name	Johny
Last Name	Ferguson
Address	Arlington St. Suite 505
Zip Code	3053217584
State	Jeffss#gmil.com
City	NULL
Phone	883-385-722

2. Disparate Data:

Having multiple systems of entry such as the CRM or the ERP is the leading cause of poor data quality and makes for a disjointed, often inaccurate perception of your customer. Even if the ERP and the CRM were integrated, you'd still have to deal with duplicates, compatibility issues (for instance, the ERP not having support for social media accounts) and integration roadblocks.

The Golden Record is created by pulling together incomplete data about all the different information fields from the systems in which they were entered. It takes information from the ERP, the CRM and any other data source to give a complete record of the entity. But this would not be possible if you have customer data streaming in from multiple source, each with its set of problems (for instance duplicates in each of these sources)

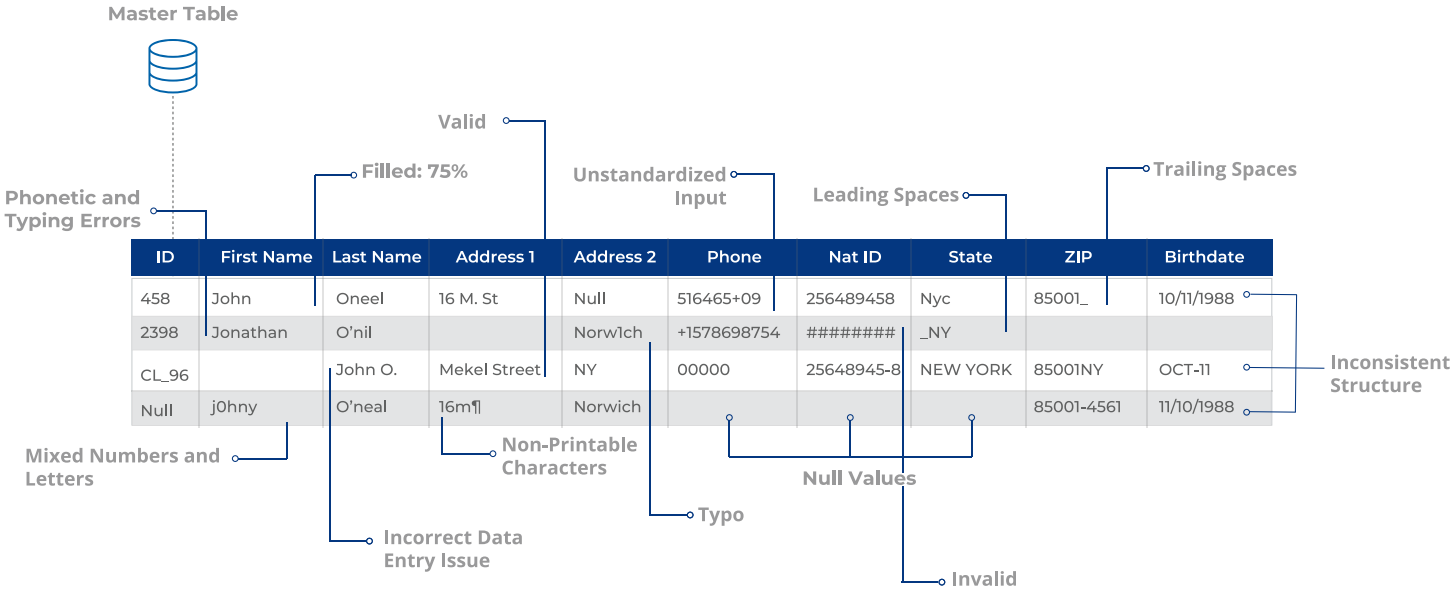


3. Dirty Data:

Any data that has not been treated is dirty data. For instance, every time your sales rep is manually typing in data, they are increasing the chances of errors. Or, every time your audience is filling in a form, it's bound to be incomplete or inaccurate. You will need to run this data through a quality check framework, where you'll 'clean' it up, to make it usable.

Dirty data remains a significant challenge and companies often get stuck at this point. You can clean up a few hundred rows of data with basic issues on Excel or using outdated ETL methods, but how do you clean up millions of rows of data? How do you ensure that your data is not affected by nicknames and missing values? How do you, 'see' the quality of your data?

Dirty, disparate, duplicate data impacts compliance, customer experience and often turns into a management issue. Therefore, the first step to getting the golden record is fixing bad data.



Resolving Data Quality Challenges & Getting the Golden Record

Historically, businesses attempted to create golden records using a combination of ETL functions, lots of complex programming, and codes to find duplicates and merge records or relied on Excel to consolidate user data.

None of these methods were able to provide a satisfying result without costing companies months of invaluable time and resources.

It would take months just to extra data, clean it up, and standardize it. Then it would take additional months to match this data and weed out duplicates.

There seemed to be no end to the work resources would have to do just to get a consolidated view of the customer – hence, most companies would either outsource the project, risking data security or they would completely skip it.

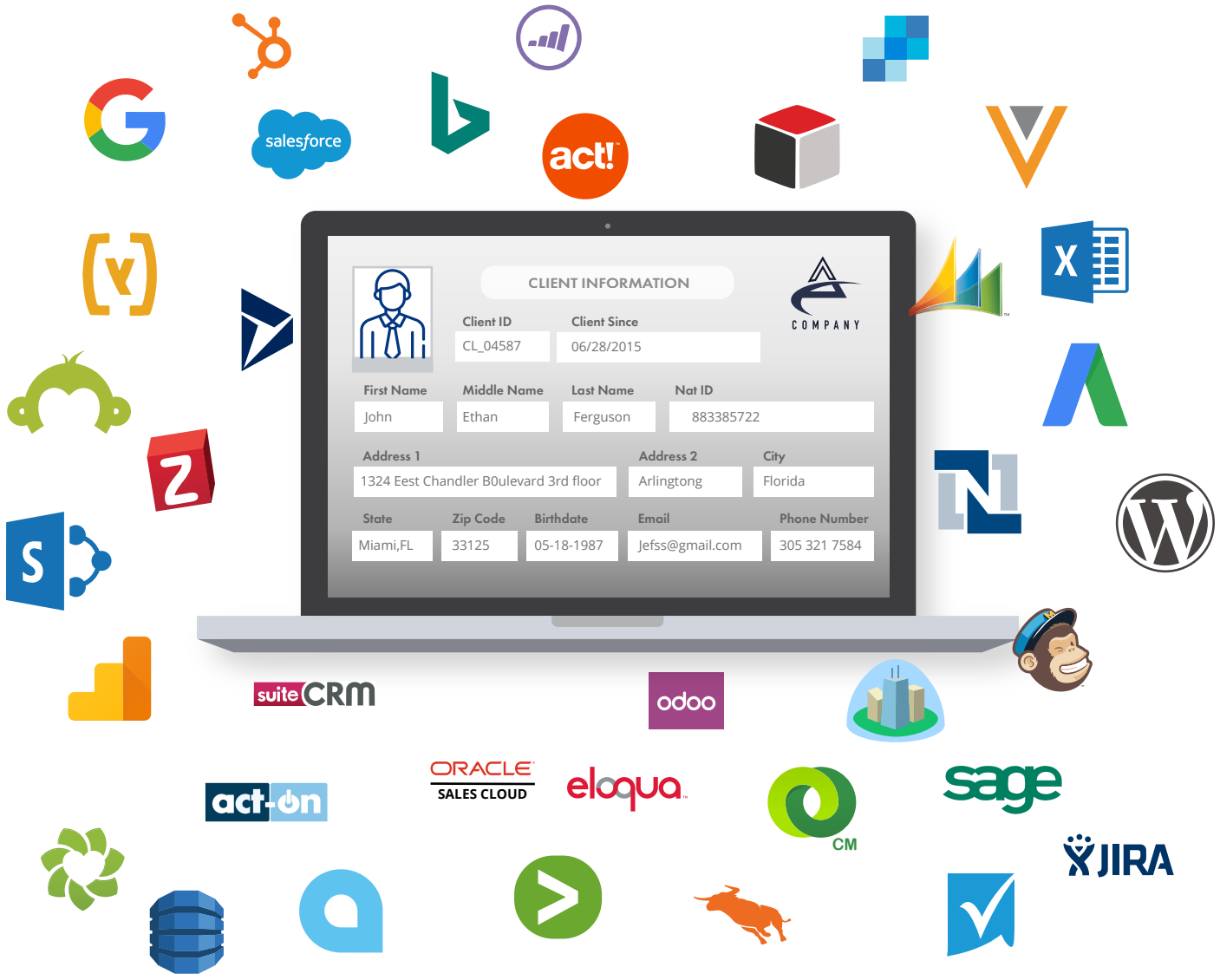
This operational inefficiency was what drove the need for data quality tools like Data Ladder's DataMatch Enterprise to be equipped with self-service merge purge features. A new class of data quality solutions, DME can be used by business managers as well as IT experts, reducing the reliance business departments have on IT departments to fix data problems.

DME has been used by government institutions such as the Department of Education and Department of Labor, as well as by Fortune 500 companies like Deloitte and HP to match, dedupe, merge/purge and create a single customer view from mailing lists and CRM data.

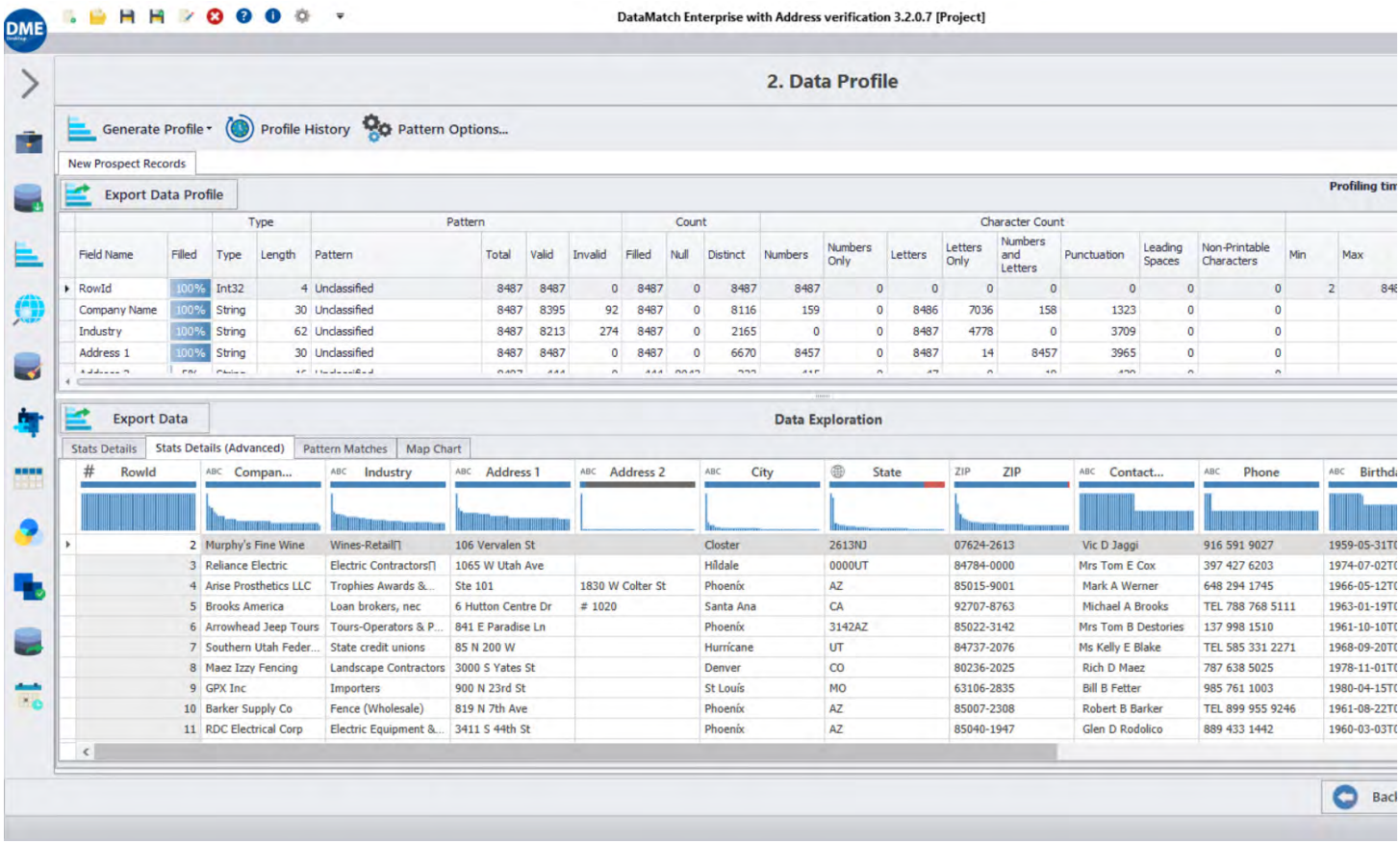
Here's a step-by-step guide on how you can use DataMatch Enterprise to merge, purge and create the Golden Record of your customer data.

Step 1: Integrate Data: Earlier on, if you had to create Golden Records, you would have to extract data from multiple sources and manually copy/paste (if there are compatibility issues) or import them into a spreadsheet file to work on them. In other cases, you would have to reach out to IT with your data and let them do the task according to their schedule.

With DME, you can integrate data from over 500 sources on to one platform. This means, your Salesforce or HubSpot data, ERP data, spreadsheet data, email or even social media data can all be integrated in the platform, saving you the hassle of extracting and uploading multiple data sheets.

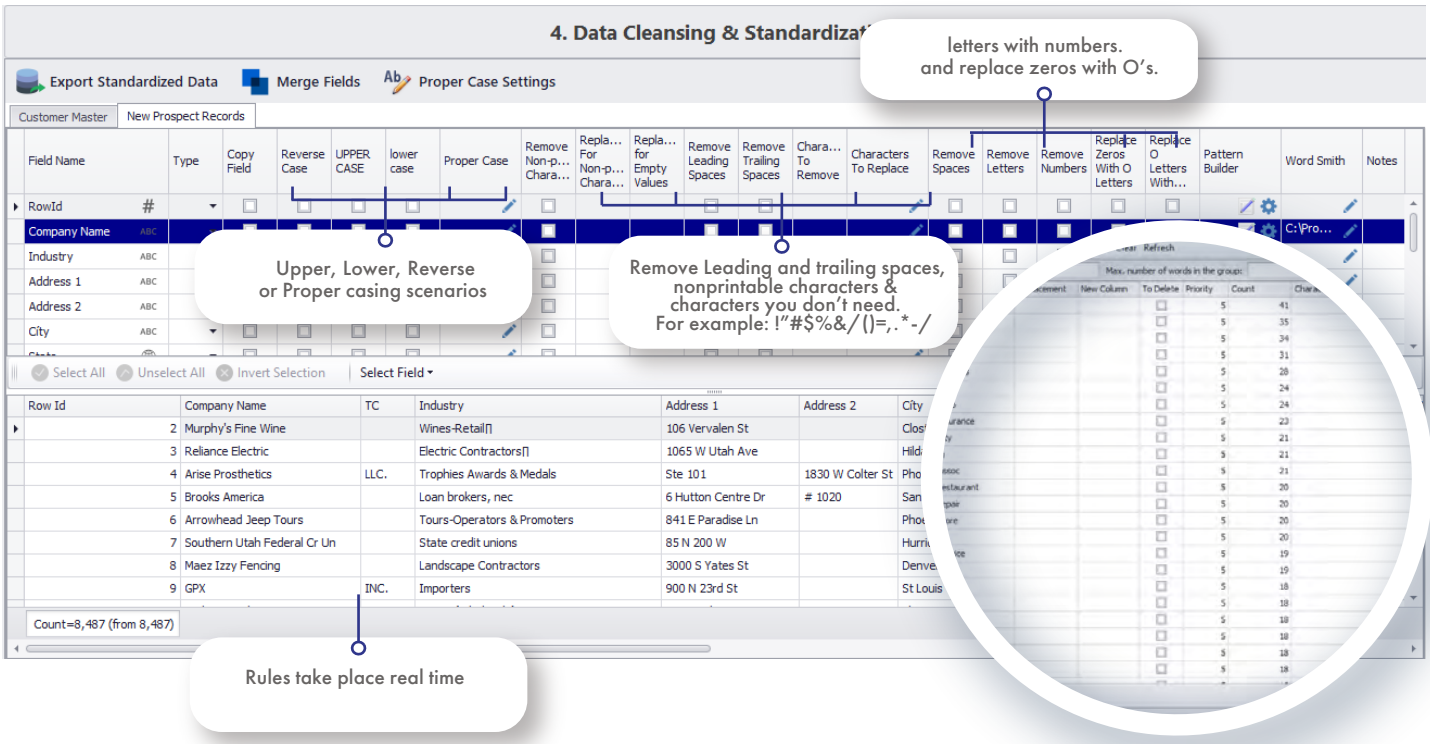


Step 2: Assess Data Quality: Once you integrate data, you can assess the quality of each data source. DME's Data Profiling feature allows you to determine the health of each column based on issues like missing values, inconsistent formats, typos and so on. It also has built-in transformations providing confidence scores, pattern analysis, visual presentation of corrupt data, mapping of data and 19 business expressions that can be used to profile data.



Step 3: Clean & Standardize Data: DME provides data cleansing and transformations that include changes to upper/lower case text, replacement of unwanted characters, non-printable characters and empty values, removal of leading spaces, trail spaces, characters and many more. Data cleansing and standardization is an essential step that will ease the process of merging and purging duplicate records.

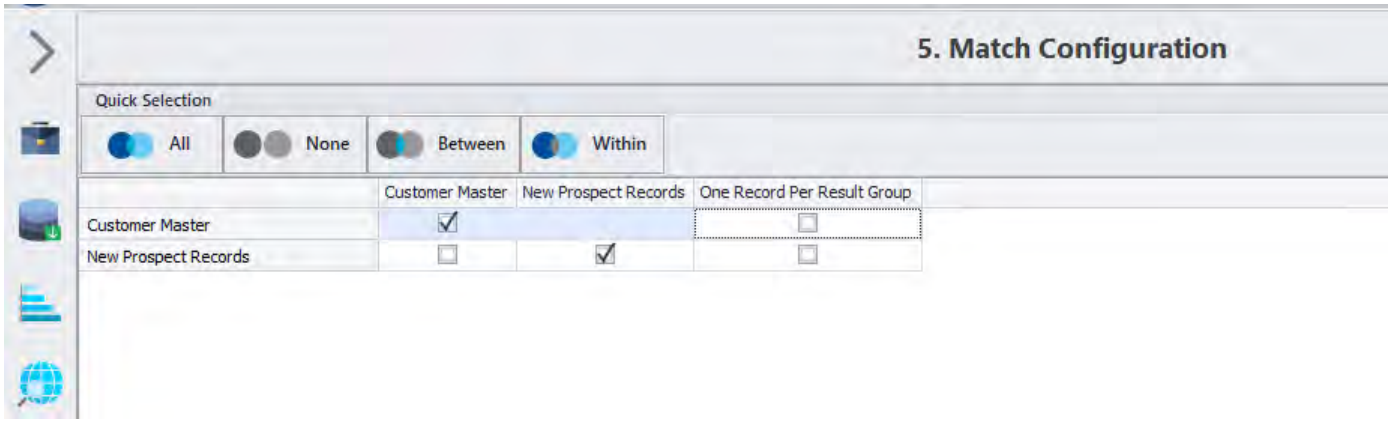
Users can also use pattern building options to build business rules catering to the unique nature of the data. For instance, users can use the WordSmith tool – a DME unique feature, that identifies nicknames from real names and helps standardize data across data sets.



Step 4: Match to Remove Duplicates: When your data and fields are standardised, look for slight variations in spelling or presentation of the data. Do you have multiple entries for John Doe, perhaps Johnny Doe or, maybe J.Doe? Beyond their name, is the rest of their information the same? If so, is it safe to assume that these entries refer to the same customer? Performing this stage will help you discern whether your brand is especially popular with women named Katie Smith who all happen to live in the same area, or whether you have multiple entries for one customer.

DME allows the matching of data in three different ways:

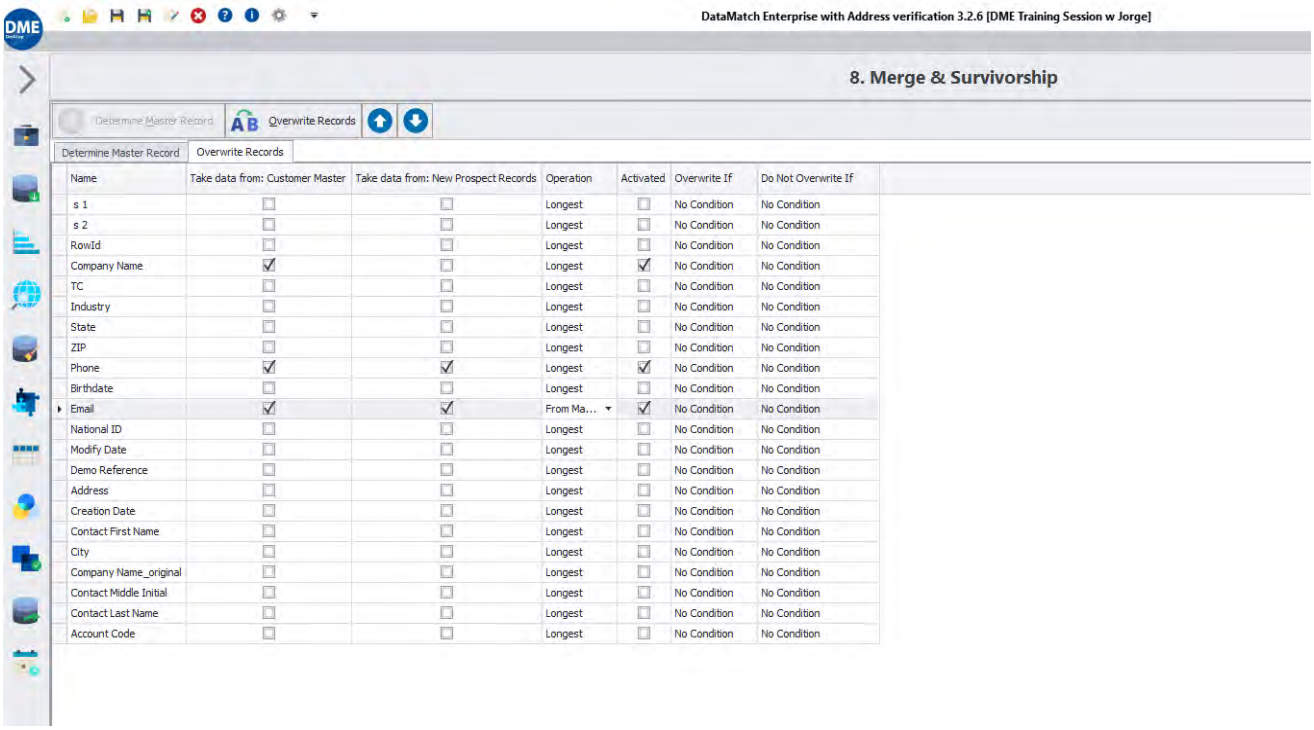
1. All - Look for matches between each data source (matches) as well as will look for matches within each data source (duplicates)
2. Between - Look for matches between the data sources and not search for duplicates within the individual data sources.
3. Within - Search for duplicates within the data sources only. No inter-data source matches will be searched.



If you've got multiple data sources and need to perform matches on more than one source at a time, these features will allow you to perform them saving on considerable time and effort.

Step 5: Merge & Survivorship: Once you're done with the cleaning, standardizing, and removing duplicates, you can use the final, 'Merge & Survivorship' option to create a master record.

This option is hugely beneficial when you will want to update your ERP or CRM data with an accurate version of the information. For instance, you can now update the customer's address across all your data sources using the Overwrite Records feature.



Step 6 - Data Enrichment: Companies aiming customer personalization goals often need to append third-party data into their records to get in-depth insights on their customers. Data enrichment is the process of augmenting this third-party to your existing database. DataMatch Enterprise has a unique **Data survivorship and enrichment** feature that lets the user overwrite records and determine the final master record *without* losing previous records.

Step 7 – Final Export: DME's Export feature will let you deal with duplicates by exporting all of the records from your data sources and create a column for the match group ID in order to "flag" the matches/duplicates. Additionally, you can also suppress (delete) the duplicates and export a clean record.

Golden records help organizations serve the purpose of delivering innovative customer service, customer management and customer personalization goals. It helps marketing and sales teams carry out data and target-driven campaigns which makes it all the more important to modern businesses today.

To get the Golden Record though, companies must focus on the consolidation of data.. For many companies, disparate, dirty data remains a significant challenge and one that prevents them from achieving their data goals. A consolidated version is the outcome of resolving those challenges, which then leads to the possibility of creating the Golden Record.

Conclusion

We highly recommend you start the journey by understanding the quality of your data and take the necessary steps to clean, standardize and dedupe it. Once you have reliable data sources to work with, the next step comes the merging/purging step. And when you conquer this step, you can then finally move to data enrichment and add value to your data – in the midst of all this, you're already creating the Golden Record!

Want to know how you can consolidate your data? Speak to our solution architect today and let us walk you through a quick demo, where we'll show you how DataMatch Enterprise can be used by business users to consolidate CRM data.

Book a Demonstration

ABOUT US

Data Ladder is a data quality software company dedicated to helping business users get the most out of their data through data matching, profiling, deduplication, and enrichment tools. Whether it's matching millions of records through our fuzzy matching algorithms, or transforming complex product data through semantic technology, Data Ladder's data quality tools provide a superior level of service unmatched in the industry.

DataMatch Enterprise, was proven to find approximately 5-12% more matches than leading software companies IBM and SAS in 15 different studies.

- Unparalleled Matching Accuracy and Speed For Enterprise Level Data.
- Cleansing beating IBM and SAS.
- Proprietary Matching Algorithms with a high level of matching accuracy at blazing fast speeds on Desktop/Laptop Hardware.
- Big Data Capability with data sets up to 100 Million Records.
- Deduplication and Merge Purge within and across any number of files.
- Suppression of existing customers or Do Not Contact from marketing lists.
- Advanced record linking technology to create data warehouses.
- Quick Data Profile tool finds and fixes Data Quality issues within the first 5 minutes of setup to improve match quality.

Free Download