



Comparison Sheet

Data Ladder Vs. SAS Data Quality



68 Bridge, St. Suite 307
Suffield, CT 06708



+1 888-779-6578



Sales@DataLadder.com



www.DataLadder.com

Table of Content

Introduction	3
1. Feature Comparisons	4
- Data Import & Integration	6
- Data Profiling	7
- Address Verification	10
- Data Cleansing and Standardization	11
- Data Matching	12
2. Data Ladder Vs SAS Data Quality	14
3. User Reviews	16
4. Comparison Chart	19
5. Conclusion	20

Introduction

With data permeating every aspect of the digital era, its quality matters more than ever, impacting business profits, customer experience, and employee productivity across the board. Businesses are realizing the costs of bad data the hard way by experiencing its effects, but on a national scale, IBM **estimates** that poor quality data costs the US \$1.3 trillion per year!

The number makes one consider data quality in a whole new light. And not just the cost, but the opportunity that lies in improving data quality for businesses. Most of the issues in data quality are a result of the right tools not being used, lack of data governance processes, and numerous **“hidden data factories”** spread across the enterprise.

As businesses see the competitive advantage in improving their data, the race for faster and more accurate is on! Rated highly in independent studies for enterprise-grade data quality and record linkage, we will compare SAS Data Quality and Data Ladder’s DataMatch Enterprise to help businesses build a solid foundation of high-quality data.

Read on to find out which platform is:

- Faster
- More Accurate
- More Affordable
- More Feature-Rich
- Easier to Use
- And Provides a Holistic Data Quality Solution for the Enterprise

Let’s get started!

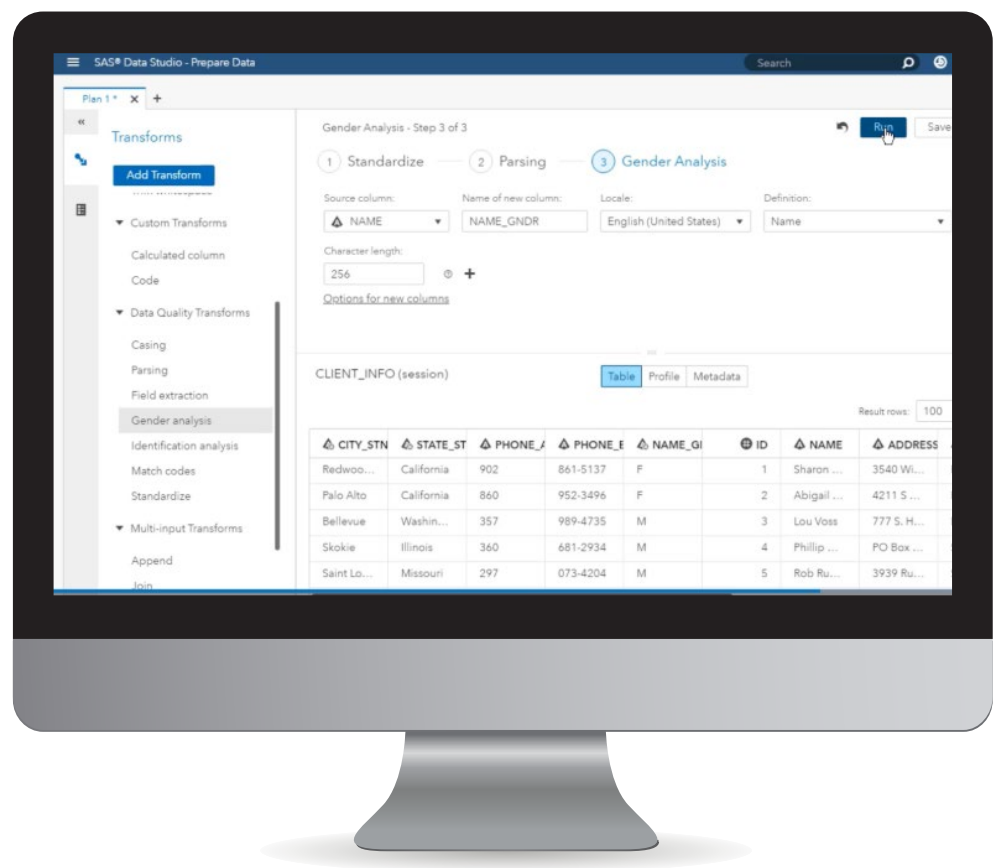
Feature Comparison: Data Ladder Vs. SAS

Data Ladder is the only company offering data matching, data cleansing, data standardization, data standardization, data merging & survivorship and data governance abilities in one single software.

SAS, on the other hand, is an enterprise that offers data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics amongst many other things. Data quality is just part of the larger picture, while data matching is part of data quality operations.

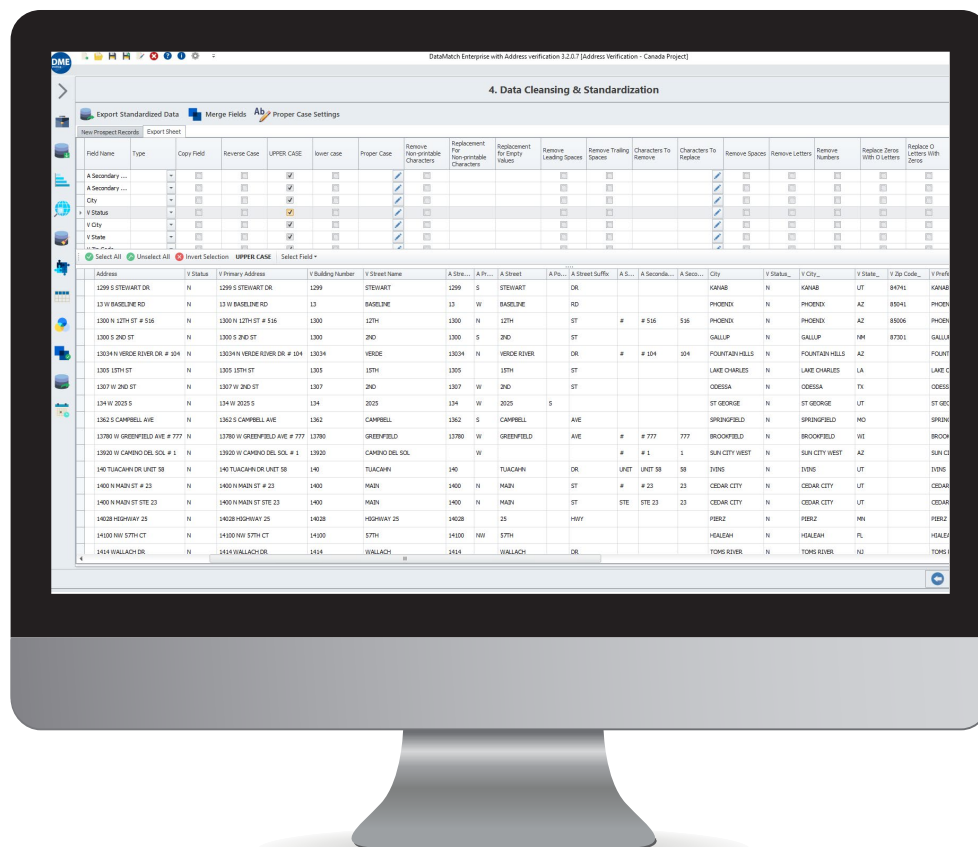
This review is, therefore, based on the data quality tools that is used by both organizations. SAS offers the following operations as part of its data quality process:

- Identification Analysis
- Locale Analysis
- Gender Analysis
- Parsing
- Extraction
- Standardization
- Casing
- Matching
- Pattern Analysis



In contrast, Data Ladder's DataMatch Enterprise Offers:

- Data Import and Integration with 150+ data sources including databases, emails, social media platforms, CRMs and text files.
- Data profiling that takes into account textual and numerical issues.
- Data cleansing and standardization that allows the user to clean and standardize their source data according to a range of preferences.
- Data matching that has a record 96% accuracy match.
- Data merge and survivorship.



We will quickly take a look at some of the main features offered by both tools and see which of them is more feature-rich.

Data Import & Integration



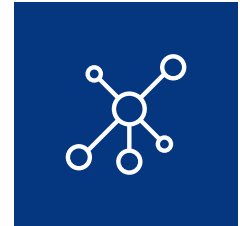
Databases



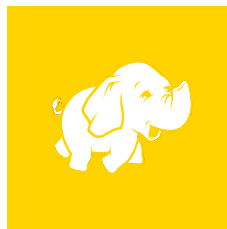
Files



CRM



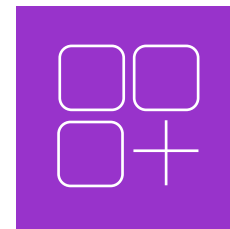
Social Media



Big Data



Email & Services



Enterprise Applications

An average company uses 64+ apps which means companies have to cope with data streaming in from multiple data sources. Data Ladder is the only service provider that gives companies an on-premises solution as well as direct integration to over 150+ data sources.

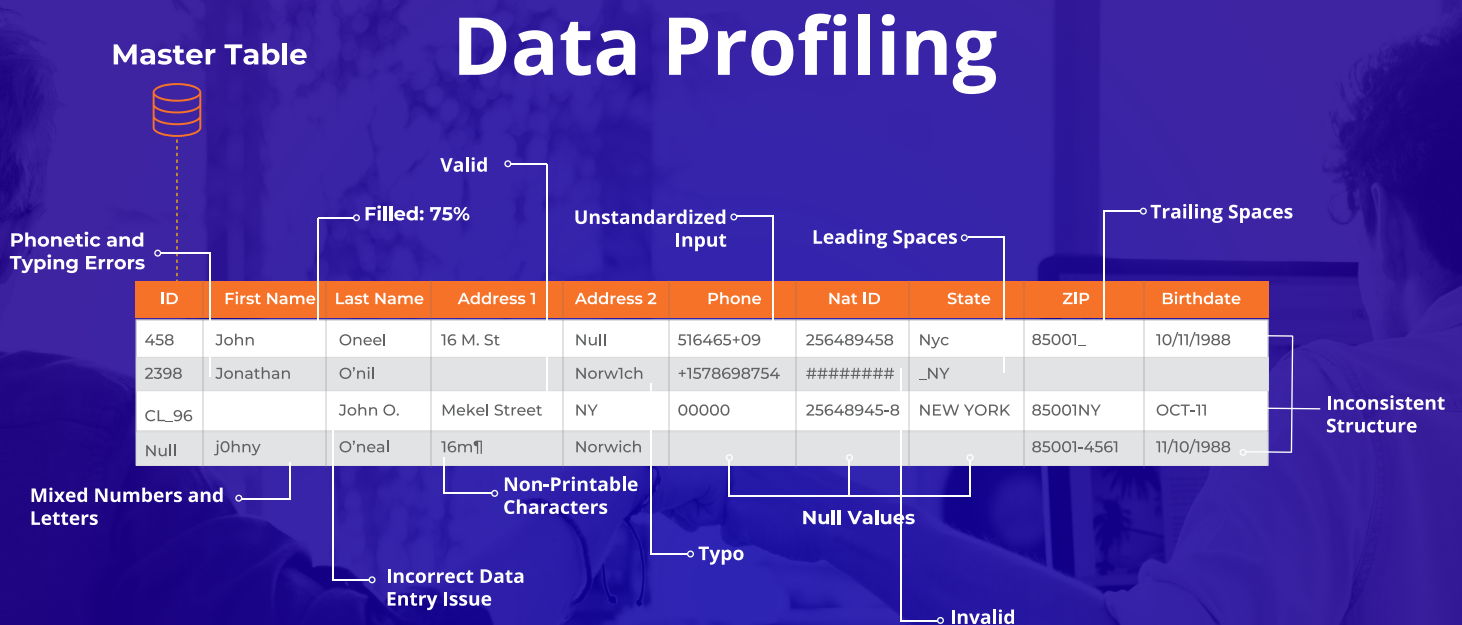
DME allows you to import data from virtually any data source, including but not limited to:

- Files (Excel, Delimited Text File, Excel (97-2003), Fixed Width Text File)
- Databases (Direct integration with SQL Server, Oracle, Teradata, and many more)
- CRM (Integration with Salesforce, MS Dynamics CRM, Sugar CRM)
- Social (Integration with Facebook and Twitter)
- Services (XML and JSON)
- Other (ODBC or Email)

SAS has three options to bring data into the environment. These are:

- **Available Tab:** Displays all tables and files that have been loaded to memory from any CAS server to which you have access.
- **Data Sources Tab:** Create connections to a database server such as a HADOOP Distributed File System (HDFS).
- **Import Tab:** Easiest option to copy documents that include local files, social media content or **Environmental Systems Research Institute (ESRI) data** (demographic and business data). You can import data from a Facebook fan page, a Twitter feed, a YouTube Account and Google Analytics. You can also import data from Google Drive.

All of these operations in SAS require technical knowledge. It is not an easy or smooth process to integrate data sources in SAS as compared to DME, where integration can be done with just a few clicks.



A professional data quality solution like DataMatch Enterprise and SAS Data Quality will always have a data profiling feature as the first step of many to help you identify inconsistencies in your data. The primary goal of data profiling is to ensure that the information in your list has the right data structure. For example, phone numbers are profiled to see if they have text or punctuation issues. Similarly, addresses are profiled to see if they have complete information including zip codes.

Before you attempt to perform any data cleansing or optimization activities, it is imperative to know the kind of problems plaguing your data. This is achieved through the data profiling feature.

Both DataMatch Enterprise and SAS have data profiling options but there are some key differences.

DataMatch Enterprise - Profiling:

1. Identify Problems: The profiling will let you discover information about your lists according to metrics as valid, invalid, filled, null, distinct, numbers, numbers only, letters only, numbers and letters, punctuation, leading spaces, non-printable characters, min, max, mean, median, mode, extreme. These units describe the overall issues with your master data.

2 Regular Expressions: Has a dedicated data profiling tab that has 19 'expressions' which are used to profile data. These expressions are tailor-made rules that were created after working with 4,500 clients and deriving common problems with inconsistent data. Users can also create their own regular expression.

3. Data Accuracy Percentage: The profiling gives a confidence percentage (score) against each field. If a [name] field has accurate data, you get a 100% confidence score. On the other hand, if a Zip field has text or punctuation issues, you may get a lesser confidence score depending on how many fields have errors.

4. Data Exploration: This feature lets you analyze which state or city you have the most hold. Data exploration helps you get an overview of your data visually – something that you would have to use third-party software to achieve.

5. Profile History: You can see a history of all the profiling you've done over your data. These are automatically saved so you will never lose any previous data or changes made to it.

Additionally, DME also has a dictionary for nicknames that identifies common nicknames. For example, Margaret is often written as Peggy or William as Wills. This is pretty helpful if you've have data from social media where the target audience tend to use nicknames instead of actual names.

SAS Data Profiling

SAS too has a Profile tab that gives you a confidence score on defined columns like the Address, Account Type etc. It doesn't, however, have the option of building your own regular expression or pattern. It also doesn't let you know whether your fields have punctuation or textual problems.

You can further profile data based on three features that lets you sort your data according to data type, location and gender.

1. Identification Analysis: This feature lets you know the types of data that you have. Identification analysis helps you understand the data by naming or identifying the type of content in each field. This is similar to DME's data profiling, but it acts as a standalone feature rather than part of overall profiling.

2. Locale Guessing: This feature simply lets you know the locale of your data using phone numbers as input data.

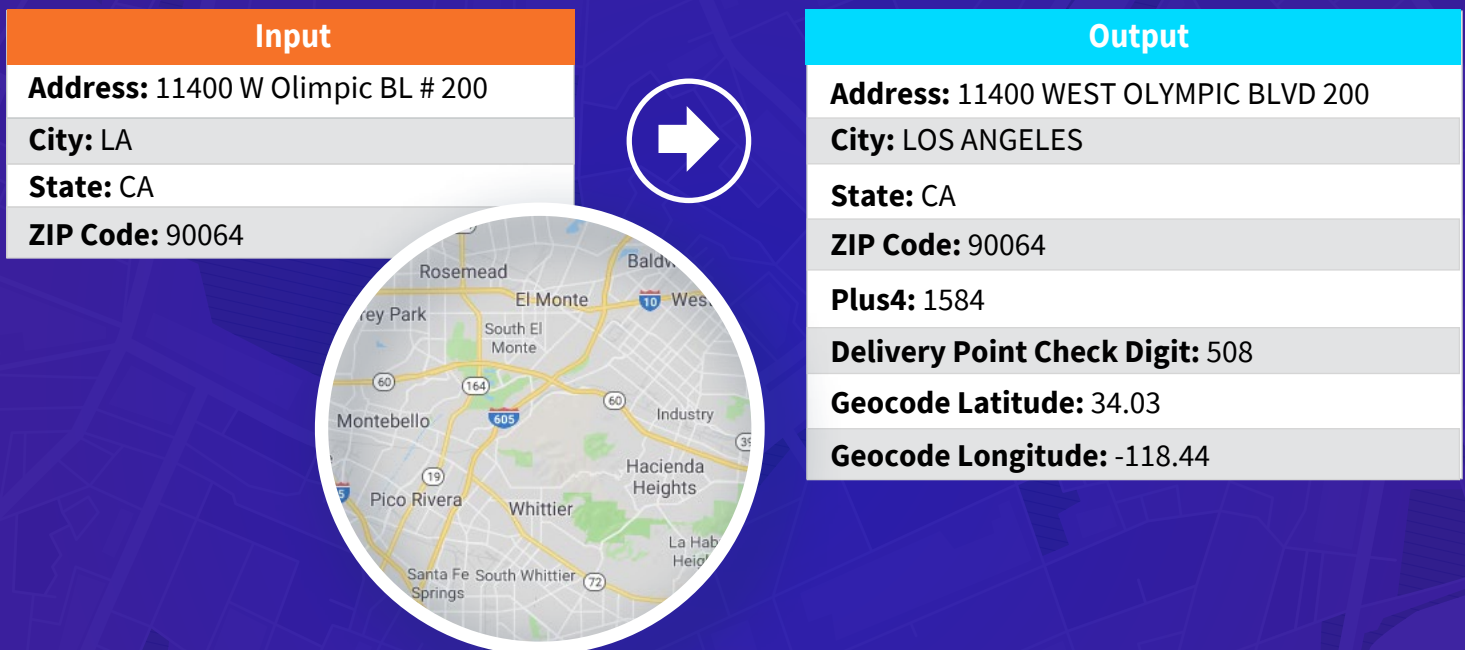
3. Gender Analysis: Using the name input, this feature gives a M/F or U (unknown) to determine the gender of contacts.

For any other customized business rules, the user will have to write a code in the language coding platform.

Address Verification

Address verification is one of the most critical components of a data quality solution. The [address] field suffers the most from inaccurate data as it's always subjugated to abbreviations, missing ZIP codes, incorrectly spelled locale names and so on. More importantly, businesses seldom match their address data against a national address database like the USPS or the UK Postal Code which means addresses don't follow national conventions for most part.

DME has a dedicated, CASS Certified address verification system that lets you match your address list against a reliable source like the USPS, the Canada Post and any other government address database. You can easily add the address dictionary of your country and the software matches your current addresses against this database, helping you verify which of your contacts have outdated or incorrect addresses.



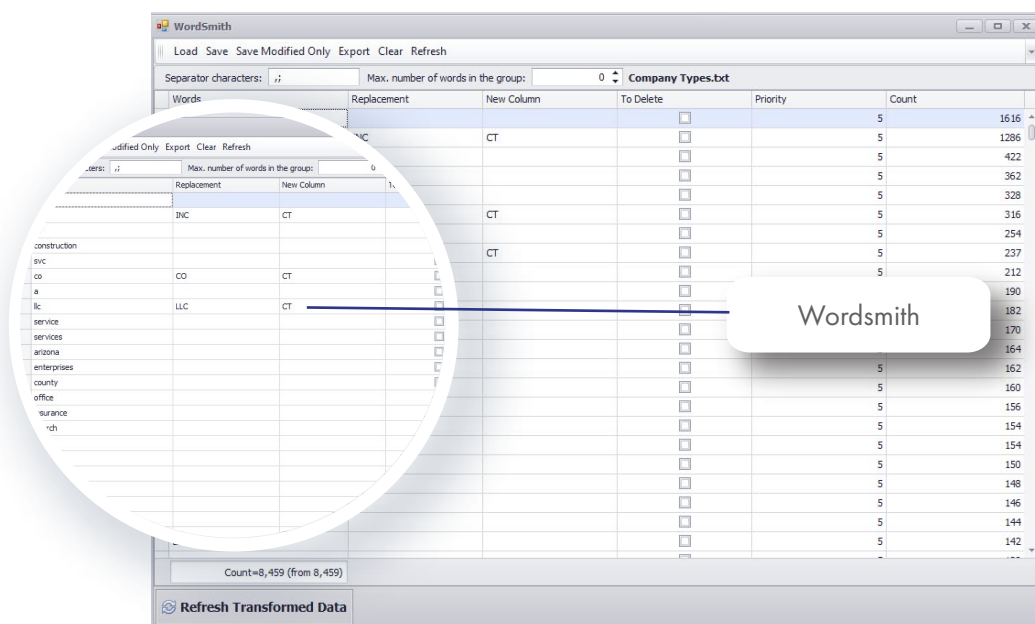
SAS does not have a dedicated Address Verification feature in its Data Quality operations. Instead, it automatically matches addresses according to the Address Verification (US/Canada) node that uses US/Canada verified address data types to replace a selected field.

Data Cleansing and Standardization

This is an important step in data quality operation. Data fields with inaccurate spellings, case settings, and other structural problems need to be cleansed by fixing the inconsistencies. Ideally, this needs to be done right within the data source so that errors are nipped in the bud. This also ensures that data across the source follow this standardization.

- Change cases and ensure that each field follows a defined standardization.
- Find and replace different characters.
- Choose whether you want to preserve abbreviations.
- Set your own exceptions.
- Choose delimiter that separates your fields such as tabs, comma, semicolons etc.
- Remove spaces, numbers, letters, non-printable characters, zeros and zeroes with letters.

WordSmith is **Data Ladder's unique tool** that allows the user to standardize the name and addresses in your lists, letting you replace all instances of that particular word with a new one. So for example, if the name, 'John' is written as jon, you can use WordSmith to replace all instances of jon with John. This way you standardize all instances of flawed data



SAS does standardization differently. It uses pattern-based rules to perform transformations and if the user uses the SAS Data Quality to perform matching operations, they do not need to pre-process their data with standardization. SAS doesn't give much room for performing customized standardization on your data and limits itself to casing and abbreviations.

Data Matching

Both the DME and the SAS Data Studio have data matching features to perform various data cleansing and data integration operations. The key difference between the two is that SAS relies **only on fuzzy logic** to match lists. It doesn't identify phonetic differences or match unstructured data. While DME's fundamental offering is data matching, SAS offers matching as a component of an overall data quality operation. As such, SAS's scope in data matching is limited.

Data Ladder's fundamental USP is the ability to use a combination of matching algorithms that cater to phonetic and textual matching that fuzzy matching fails to do. In fact, Data Ladder has a proprietary algorithm that is used to deliver 96% data matching accuracy. This figure was proved in a comprehensive record linkage study conducted by Curtin University.

The results proved that Data Ladder had a 96% matching accuracy as compared to IBM and SAS which stood at 85% and 83% respectively.

With more than 4000 matching and cleansing projects performed over a decade, Data Ladder has mastered the science of accurate data matching – a feat that data quality solutions have yet to offer.

Additionally, Data Ladder offers the ability to perform cross-jurisdictional matching where users can match multiple data sources instead of lists within a specific data source. Cross-jurisdictional matching is an important feature that government and public sector organizations need to create research reports, receive grants and optimize government processes.

Data Ladder vs. SAS Data Quality

DataMatch Enterprise

Data matching at 96% accuracy for 40K records.

Offers data matching as a key feature along with other data quality operations.

Highly user-friendly, requiring no additional technical knowledge.

Offers everything in one central, on-premises platform.

Has predefined rules based on experience working with 4500 clients which cover almost every major data quality issue.

Has a special WordSmith tool for parsing and standardization that is not offered by any other data quality solution.

Has a CASS certified Address Verification system that gives the user flexibility to add government databases and dictionaries for address verification.

Does not require any additional licensing apart from the one-time purchase license.

Is designed for businesses with a focus on data matching, data cleansing, and data standardization.

Pricing the lowest in the industry, 90% less than SAS.

SAS Data Quality

Data matching at 84% accuracy for 40K records.

Offers data matching as a subset of data quality operations.

Requires technical knowledge and SAS language to create customized rule.

Has multiple platforms for multiple data management tasks which could become confusing to use.

Has very limited pre-defined rules and requires SAS language expertise to create custom business rules.

Has a special WordSmith tool for parsing and standardization that is not offered by any other data quality solution.

Has a CASS certified Address Verification system that gives the user flexibility to add government databases and dictionaries for address verification.

Does not require any additional licensing apart from the one-time purchase license.

Is designed for businesses with a focus on data matching, data cleansing, and data standardization.

Pricing the lowest in the industry, 90% less than SAS.

Features of the Solution	Data Ladder	IBM Quality Stage	SAS Dataflux	In-House Solutions
Match Accuracy (Between 40K to 8M record samples)	96%	91%	84%	65-85%
Software Speed	Very Fast	Fast	Fast	Slow
Purchase/ Licensing Costs	80-95% below the competition	\$370K+	\$220K+	\$250K+
Time to First Result	15 Minutes	2 Months+	2 Months+	3 Months+
Comments	Above tests were completed on 15 different product comparisons with university, government, and private companies (80K to 8 million records). This includes the effect of false positives.	Need multi-threaded, in memory, no-sql processing to optimize for speed and accuracy. Speed is important, the more match iterations you can run, the more accurate your results will be.	Include base license costs, 2014 process or later, in-house, includes salary + benefits. Not in-house implementations had a 10% chance of losing in-house implementations had lost the core member who ran and understood the matching program.	A metric for ease of use. This is the time of first result, not necessarily full cleansing.

Making the Choice Between Data Ladder and SAS

SAS Data Quality originally started off as a separately sold product from DataFlux, a subsidiary of SAS. Even today, data veterans continue to refer to the product as DataFlux. Over the years, the platform evolved to provide a more complete solution to manage the entire data quality life cycle. It does this by allowing users to:

- Integrate structured data from multiple sources
- Profile it to understand hidden relationships
- Cleanse it by correcting non-standard and duplicate records
- Perform entity resolution using established data matching algorithms
- Build a Master Data Management (MDM) foundation
- Visualize lineage to see how changes affect other assets

Reviewers claim that SAS Data Quality does the job well - if you're a dedicated programmer processing massive datasets to build a Master Data Management (MDM) hub.

On the other hand:

Data Ladder's data quality products are designed to help business users get the most out of their data through data matching, profiling, deduplication, and enrichment tools. Whether it's matching hundreds of millions of records through our proprietary 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. The company enables users to:

- Integrate multiple sources simultaneously from virtually any source system.
- Profile their data within a couple of minutes to visually show data quality issues.
- Clean it using point-and-click actions to transform their data visually.
- Standardize it by creating custom libraries or using the 300,000+ pre-built rules.
- Verify addresses and enrich existing records with geocoding data.
- Deduplicate their data with a combination of proprietary and established matching algorithms.
- Link records across the enterprise for entity resolution and data enrichment purposes using proprietary fuzzy matching technology, rated as the fastest and most accurate.
- Handle both structured and unstructured data with ease.
- Leverage advanced semantic-based machine learning technology to process product data.
- Deploy a data quality firewall to catch data quality issues before they make it into your systems.

Reviewers claim that Data Ladder's tools are easy to use and offer unparalleled efficiency and productivity increases when the use was comparing disparate data sets, cleansing them, and implementing clean-data initiatives across the enterprise.

"DataMatch Enterprise, Powerful and Easy to Use"

It does a great job with data cleansing making the matching process even more powerful and being able to merge rows, with very flexible rules for the final export is extremely helpful.

SAS and Data Ladder User Reviews

Ease of Use

While **SAS Data Quality** is highly acclaimed for foundational master data management, ease of use is not its strong suit, as evident from user responses on trusted software review websites:

"SAS is not something you can easily pick up. It requires some background knowledge on what each procedure does (ex: reg, glm, freq, means, genmod)."

"SAS is not user-friendly and I had to seek assistance on multiple occasions. You have to know programming code and remember it for analysis."

"The interface is extremely difficult, and the programming language is not at all intuitive."

"Cannot "see" the data to look for abnormalities without running a proc freq. Having access to the visual interface allows for scanning the data."

Issues that users face most often with SAS Data Quality include:

- Difficult to understand user interface.
- The requirement of programming knowledge.
- Poor documentation available.
- Lack of useful guidance from support staff.
- No way to visually see the data as it changes.
- Not built for the business user.

Data Ladder's biggest selling point is its ease of use, designed specifically for both business and IT users. Customers who have never used a data management tool before are able to create data quality projects easily within the modern, intuitive interface with basic training, as is evident from reviews:

"DataMatch Enterprise is easy to learn and use. It's easy to review results. Saves us tons of time in manually checking records."

"I like that it is very straightforward and easy to use. With very little training, we had it up and running."

"This program is far more effective, user-friendly and reasonable than many other data cleansing programs I checked out prior to making my purchasing decision and their customer service is really great."

"I like the ease of use for this program when matching data from Access tables to confirm how successful our business practices are."

When using Data Ladder software, customers especially liked the following:

- The modern, highly intuitive, code-free interface.
- Extremely responsive and helpful support.
- Flexible creation of data quality rules for both business and IT users.
- Very easy to review result and see their data as it transforms.
- Straightforward setup - very little training required to get started.
- The ease of integrated data sources across the enterprise.

Quick Comparison Chart

Product Capabilities	Data Ladder	SAS Data Quality
Enterprise-grade Data Integration	✓	X
Data validation against statistical measures	✓	✓
Uncover relationships across tables, databases, and enterprise applications	✓	✓
Establish trends and commonalities	✓	✓
Profile massive datasets in minutes	✓	X
Identify duplicate records and match linked records across the enterprise	96% Accuracy	85% Accuracy
Advanced domain-specific libraries such as nicknames, addresses, phone numbers	✓	X
Create custom standardization rules	✓	✓
Entity resolution using a fusion of established (Probabilistic Jaccard, Levenshtein distance, etc.) and proprietary fuzzy matching algorithms	✓	X
Scoring and cut-off modes, predicates during data matching process to eliminate false positives	✓	X
Multi-table selection, semantic equivalence classes	✓	X
Multi-language support	✓	X
Pattern matching to intelligently search and parse out data	✓	✓
Advanced filtering using wildcards, and/or, or/not statements	✓	✓
Reusable workflows and visual orchestration	✓	✓
In-memory processing architecture	✓	✓
Machine learning-enabled	✓	X

Contextual recognition capabilities to make sense of unstructured data	✓	X
Unstructured data classification and organization (UNSPSC, ICD-10, etc.)	✓	✓
Taxonomy development	✓	✓
Establish trends and commonalities	✓	✓
Purchasing and Licensing	80-95% Less than Comparable Solutions	\$220,000+

Conclusion

The choice between SAS Data Quality vs Data Ladder's DataMatch Enterprise is one that is dependent upon your business requirement in terms of data initiatives. If you're looking for a solution that offers a low-cost, high-impact data cleansing and matching, then DME is the optimal choice. On the other hand, if you're looking for advanced data solutions like big data predictive analysis, data investigation and other complex data operations, then using SAS would be a better option.

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](#)