

# COMPARISON STUDY

## IN-HOUSE VS. BEST IN CLASS DATA MATCHING SOLUTIONS

While many companies say that they have 100% accuracy, it isn't necessarily true. Many companies develop solutions internally and often don't take the time to test their matching system. Consequently, this can lead to lower match accuracy, missing as many as 1/3 of matches.

These companies underestimate the complexity of their problems, looking at one data set during a set point in time and not realizing that more sophisticated methods could save them time and money. In-house solutions are also typically slower than expert software programs, which doesn't give users that much time to review the matches, ultimately resulting in a very expensive project that doesn't achieve the end goal.

### Making the Case for Software Solutions

So what happens when your in-house IT employee leaves? Often times, the project gets left to someone who doesn't understand it, or in many cases the data matching project gets tabled. In the meantime, the allocated project budget still exists, so money is being wasted while the business need is not being fulfilled.

There have been several cases where in-house solutions put the company at risk. One government institution needed to match between large populations for a study, so they decided to use a university to perform the data matching. Paying \$500,000 a year to have it maintained, the people who created the code have since left the university, but now the institution is finding bad matches in the data, which now impacts a \$10 million reporting project. This puts the both the project and many careers at risk.

Many corporations have also experienced this issue, finding bad matches with internal solutions, but due to an internal employee leaving, the company ends up missing many opportunities.

Another common situation is when the IT department develops a matching solution without any real knowledge of the business or problem in question. The business user comes in and is able to quickly see the profile data and perform match iterations, but when it comes time for them to submit an IT request and review the matches, there is an issue with the methodology of the matching. There isn't a way to adjust it, as it's what IT built and there's no time to create a full solution.

All of these examples show the complications of using an in-house solution. If you're managing your data quality through an in-house program, it's time to reevaluate how effective your program really is in terms of time savings, cost and ease of use.



# COMPARISON STUDY

Vendor	Match Accuracy (Between 40K to 8 million record samples)	Software Speed	Purchase/ Licensing Costs	Time to First Result
Data Ladder	96%	Very Fast	80-95% below the competition	2 Hours
IBM Quality Stage	91%	Fast	\$370K+	2 Months+
SAS Dataflux	84%	Fast	\$220K+	2 Months+
In-House	65-85%*	Slow	\$250K+	3 Months+
Notes	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.	Includes base license costs. 2014 prices or later, in-house, includes salary + benefits. Note in-house implementations had a 10% chance of losing in-house personnel, so over 5 years half of 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 to first result, not necessarily full cleansing.

*It is evident that no solution has 100% match accuracy, as seen in this chart. Benchmarked in 15 different studies, DataMatch Enterprise was proven to find, on average 5-12% more matches than leading software companies IBM and SAS.*

DataMatch Enterprise can handle many of the issues that compromise your data systems. Our system is scalable -- even with large datasets, the information can be analyzed with lightning fast response times. Our platform is a robust approach to making imperfect data usable. Making the right connections with any type of data, DataMatch Enterprise can correct many data quality issues, from spelling errors to redundancies.

## ABOUT DATA LADDER

Data Ladder, headquartered in Cambridge, Massachusetts, is a data quality and cleansing company dedicated to helping you get the most out of your data through data matching, profiling, deduplication, and enrichment. Working with clients such as Hewlett-Packard, Hoover's, and Cardinal Health, Data Ladder's suite of data cleansing services is in use across the Fortune 500 and among many mid-size companies.

Corporate Address: 68 Bridge St. Suite 304  
Suffield, CT 06078

Phone: (866) 557-8102  
info@DataLadder.com  
DataLadder.com

