Data Ladder
Get the most out of your data

DataMatch Enterprise™

Curtin Study Comparing Record Linkage Accuracy

- Record Linkage Accuracy Comparison
- Independent University Testing Based on Years of Research
- "Record Linkage Accuracy Based on Matches Found and Least False Positives"

68 Bridge, St. Suite 304
Suffield, CT 06708

+1 888-779-6578

Sales@DataLadder.com

www.DataLadder.com

Data Ladder
Get the most out of your data

Curtin University Record Linkage Accuracy Test

There are many vendors for record linkage, but it is difficult to determine who is the best as testing different vendors is time consuming, difficult, subject to interpretation along with non disclosure agreements.

As part of our ongoing research and development, Data Ladder participated in a record linkage study at Curtin University. Since linking patient data is crucial for both providing healthcare as well as conducting healthcare research (How smoking effects the population as a whole, etc) Curtin University's Healthcare college has a Centre for Data Linkage. The purpose of the Centre is to evaluate software packages' record linkage accuracy as better record linkage leads to better healthcare research and better patient outcomes.

Over several years, The Centre for Data Linkage painstakingly developed 3 data sets from 40K records to 4 million records that had been manually reviewed with correct matches already determined. Based on this data set, they evaluated 11 software vendors on record linkage accuracy. The accuracy measurement used took into account both finding all matches possible, and also adjusted scores if false positives (Bad matches) were found.

The study included participants from IBM, SAS DataFlux and more. Do note, no participants sponsored the study in any way to ensure the results were neutral.

Curtin University

Population Health Research Network
Centre for Data Linkage

Data Ladder
Get the most out of your data

DataMatch Enterprise™

Industry Leading Accuracy and Speed

Data Ladder's DataMatch Enterprise™ product won for record linkage accuracy across small, medium, and large data sets with the fastest speed and lowest cost. IBM came in second place with good match accuracy results at a much higher price point.

"Our team has been hard at work the last several years improving matching accuracy, so this is truly exciting news,"said Nathan Krol, founder and CEO of Data Ladder. "At our price point, these types of tools - once only accessible to large companies with dedicated resources - are now accessible to a wide variety of organizations."

	Match Accuracy				
	40K Records	400K Records	40K to 4M Records	Speed	Purchase / Licensing Costs
Data Ladder	96%	91%	95%	Very Fast	Low
IBM Quality Stage	88%	87%	91%	Fast	High ($250K+)
SAS DataFlux	84%	84%	81%	Fast	High ($250K+)

Note: The above tests were completed on internal test data (External confirmation in process). Take into account, these tests were done using our proprietary algorithms; no pre-processed algorithmic results were used.

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

Shelley Hahn - Business Development Associate

Precision, Recall and F-measure

Are types of measurements used to assess linkage quality (Australian Center for Data Linkage).

Precision

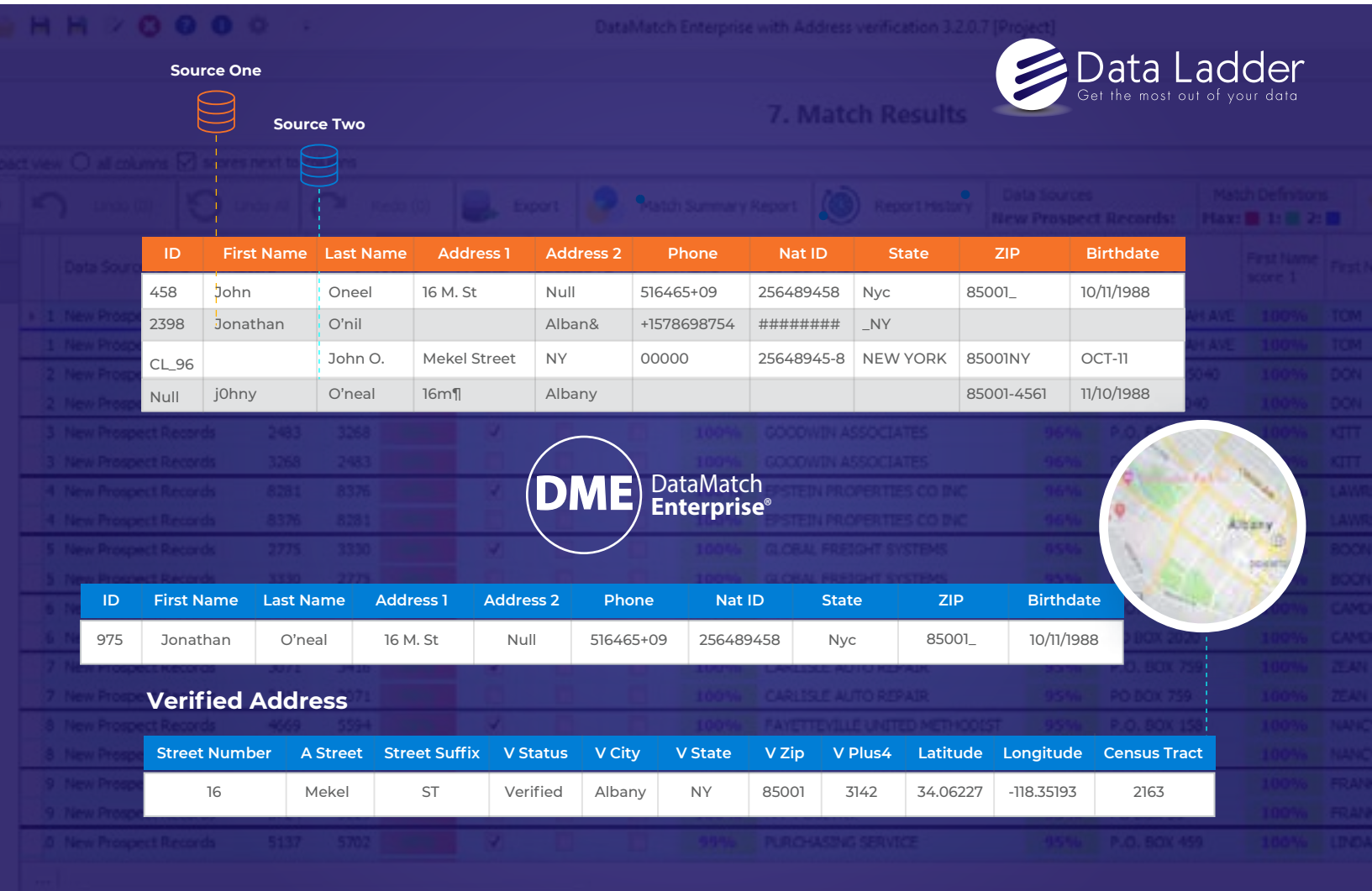
Refers to the proportion of returned matches that are true matches (sometimes referred to as positive predictive value).

Recall

Is the proportion of all true matches that have been correctly identified (Recall is also known as sensitivity).

F-measure

The f-measure is seen as a way of finding the best accommodation between Precision and Recall.



ID	First Name	Last Name	Address 1	Address 2	Phone	Nat ID	State	ZIP	Birthdate
458	John	Oneel	16 M. St	Null	516465+09	256489458	Nyc	85001_	10/11/1988
2398	Jonathan	O'nil		Alban&	+1578698754	########	_NY		
CL_96		John O.	Mekel Street	NY	00000	25648945-8	NEW YORK	85001NY	OCT-11
Null	j0hny	O'neal	16m¶	Albany				85001-4561	11/10/1988

ID	First Name	Last Name	Address 1	Address 2	Phone	Nat ID	State	ZIP	Birthdate
975	Jonathan	O'neal	16 M. St	Null	516465+09	256489458	Nyc	85001_	10/11/1988

Street Number	A Street	Street Suffix	V Status	V City	V State	V Zip	V Plus4	Latitude	Longitude	Census Tract
16	Mekel	ST	Verified	Albany	NY	85001	3142	34.06227	-118.35193	2163

Data Ladder
Get the most out of your data

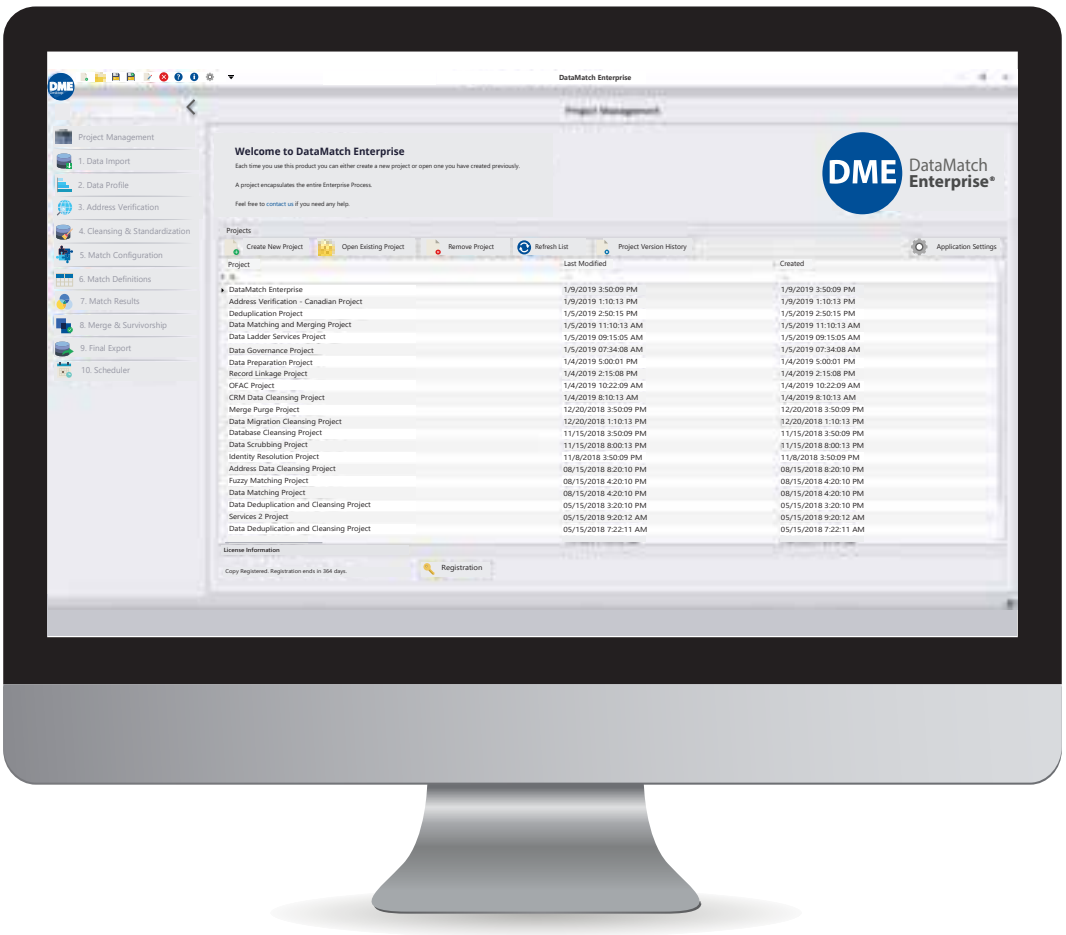
Standardization and Simplicity

Oftentimes, data is inconsistent both in formatting and in wording. Example: Road Vs Rd, Sue Vs Susan, International Business Machines vs. IBM, etc.

This is why we have included several standardization libraries with DataMatch Enterprise as well as allowing for quick and simple creation of your own libraries to be used repeatedly.

Data cleansing and fuzzy logic can be complicated. This is why we invested strongly in a world class visual interface that minimizes the number of clicks needed to complete a data cleansing and matching project.

Also, DataMatch Enterprise provides a visual graphic user interface that can be filtered and customized whichever way works best for your project.



Data Ladder
Get the most out of your data

PC Specifications

DataMatch Enterprise™ has been designed for use on Windows based PC systems. To get the best results, you need to match the hardware to the job in hand. For smaller jobs where performance is not an issue, we would normally recommend a standard desktop machine that will no doubt be well above the minimum requirements (stated below). For larger data sets, you should consider upgrading the hardware according to the recommendations below. We have identified the main areas which affect the software's performance below.

Processor Speed: Processor speed or CPU speed (measured in GHz) controls the rate at which information can be processed. The higher the CPU speed, the faster the processing.

Number of Cores: The most critical parts of the matching process in DataMatch Enterprise™ are working in parallel; having more cores means better performances.

Hard Disk Speed: The faster the hard disk the better; DataMatch Enterprise™ is disk intensive which means the rate with which the data can be read and stored from and to the disk is an important factor. The higher the RPM the greater the data transfer rate is. Deployment of data and temporary files are covered by a separate document to optimize DataMatch Enterprise ™ performance.

A SCSI interface will further increase the transfer rate from the disk thus being beneficial, especially, when processing large volumes of information (although SATA is sufficient for the great majority of organizations). Using a RAID optimized for performance can also be useful.

Our general recommendation is to use Solid State Drives (SSD). They have a proven speed superiority over traditional hard drives plus they do not suffer from regular disk fragmentation.

Memory: RAM is a key factor when processing large sets of data. Since DataMatch Enterprise ™ is NoSQL software, a sufficient amount of RAM is critical.

Operating System: DataMatch Enterprise™ is compatible with Windows operating systems from 7 onward (including server versions). We usually recommend using the most up-to-date version of a 64 bit operating system since this will benefit DataMatch Enterprise™ in terms of memory usage and robustness.

Data Ladder
Get the most out of your data

Operating System

Optimum	Minimum
4 or more cores	Dual Core
Intel Core i7	X
16 or more GB of RAM	4 GB of RAM
SSD, or SCSI HDDs	X
SSD (Free space requirements depends on the size of input data)	Hard Drive (free space requirements depends on the size of input data)
1 TB Disk Space	500 MB Disk Space
Microsoft Windows 7 SP1 and above, 64 bit (Current Generation)	Microsoft Windows Sever 2003, Windows 7, 32 bit versions

Data Ladder
Get the most out of your data

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.

Why Data Ladder

It's simple: our user-friendly and powerful software helps business users across many industries manage their data more effectively and drive their bottom line. Our powerful software suite, DataMatch Enterprise, was proven to find approximately 5-12% more matches than leading software companies IBM and SAS in 15 different studies.

Let **Data Ladder** be your partner in your next marketing campaign.
Increase your sales by offering data cleansing services through
DataMatch Enterprise™

Free Download