

- ❖ Record Linkage for Large Data Sets
- ❖ Fuzzy Matching
- ❖ Deduplication
- ❖ Merge/Purge

# DataMatch Enterprise Server Edition

**Robust Speed and Accuracy  
for the Enterprise User**



# DataMatch Enterprise Server Edition

DataMatch Enterprise Server Edition 2015 is our best in class software tool for enterprise users.

With big data capability up to 100 million records, DataMatch Enterprise recently beat IBM and SAS in an independent study with unparalleled matching accuracy and speed for enterprise data level cleansing.

Results from the independent study are shown below:

Linkage Scenario	Runtime	Linkage Quality		
		Precision	Recall	F-Measure
Deduplication 40,000 records	12 Seconds	95.3	97.1	<b>96.2</b>
Deduplication 400,000 Records	3 Minutes	92.6	93.2	<b>92.9</b>
Linkage of 400,000 records to 4 million records	45 Minutes	91.5	93.1	<b>92.3</b>

*Note the above tests were completed on internal test data (external confirmation in process). Note that these tests were done using our proprietary algorithms, no pre-processed algorithmic results were used.*

*Using a standard data set comparing 12 data cleansing vendors including IBM and SAS, DataMatch Enterprise was consistently the fastest and most accurate in terms of identifying true matches and avoiding false positives.*

# DataMatch Enterprise Server Edition

## DataMatch Enterprise Server Edition

### FEATURES

- **Advanced record linkage functionality to create data warehouses**
- **Data deduplication**
- **Mid-process cancellation**
- **Matching statistic reports**
- **Automated name parsing and data enhancement of gender, salutations, and more**
- **Information standardization across all fields**
- **Data enhancement**
- **Merge/purge functionality**
- **Suppression of data from existing customer or “Do Not Contact” lists**
- **Extraction of required information**
- **Geocoder latitude and longitude add-in available**

The latest version has been revamped to include a more visually oriented platform, which provides a more user-friendly experience, allowing the user to filter and customize the tool for their ease.

In addition to standardization libraries, DataMatch Enterprise Server Edition includes scalable configurations for deduplication and record linking, suppression, enhancement, and extraction.

Whether you need to increase response time, increase sales and conversion rates, or reduce time and costs, our server edition can meet the needs of the most large-scale business users.

# DataMatch Enterprise

## Address Standardization Tool

### FEATURES

- **Import files from over 270+ different formats**
- **Nationally CASS Certified™ - Process your data at speeds up to 2.1 million records per hour**
- **GOLD PAVE™ Certified - prepare your mail with the lowest postage rates**
- **PostalOne! MAIL.DAT – submit data electronically to the USPS for postal acceptance and payment purposes**
- **Integrated PDF creator and WinZIP compatible compression**

Check the validity and deliverability of a physical mailing address with our updated data quality technology. DataMatch Enterprise Server Edition comes equipped with address validation and geocoding technology, which helps standardize your address lists for mailings.

Standardization converts an address to a standard format by correcting the address and adding missing information (such as a zip code or a suffix). It is then compared against a list of valid addresses to determine validity.

Learn about the important features of our address standardization tool.

You also have the ability to enhance each matching address and add ZIP+4 level latitude and longitude values for the best in mapping precision.

# Sample Input/Output

Input Information	
Company	Bluebird Bookkeeping Services
Address	Suite 1200 5670 Wilshire Bl
City	Los Angeles
State	California
Zip Code	90036
Toolkit Standardized Information	
Company	Bluebird Bookkeeping Services
Address	5670 WILSHIRE BLVD STE 1200
City	LOS ANGELES
State	CA
5-Digit ZIP Code	90036
Plus 4 Code	5679
Delivery Point and Check Digit	367
Carrier Route	B045
Line of Travel No.	0239
Line of Travel Code	T
Address Type	K
Return Code	24
Errors and Warnings	D#E#M#N#V#
Building Number	5670
Pre-Direction	
Street Name	WILSHIRE
Post Direction	
Suffix	BLVD
Secondary Name	STE
Secondary Number	1200
Vacancy Flag	N
County FIPS Code	037
State FIPS Code	06
County Name	LOS ANGELES
Congressional District No.	37

With Geocoding Tool	
Latitude	34.06227
Longitude	-118.35193
Census Tract	2163
Census Block Group	1

## Includes

CASS Certified™  
 RDI™  
 ZIP+4™  
 DPV®  
 eLOT®  
 LACSLINK®  
 SuiteLINK®

# DataMatch Enterprise Server Edition

## Geocoder Latitude and Longitude

### FEATURES

- Add Latitude/Longitude coordinates to your address lists at the ZIP+4 level
- Add latitude and longitude coordinates carried out the 6th decimal
- Add street-segment points
- Provide county name and county FIPS code to the ZIP+4 level
- Calculate distance from a specific ZIP+4
- Perform range searches (e.g. between 10-20 miles from a specific ZIP+4)
- Map with Google, Bing, or Yahoo! Local Maps

Geocoding provides the geographical coordinates for a specific address or location. Whether you need to analyze intelligence data or visualize distance, there are numerous applications of this very important tool.

Our geocoding module provides exact latitude and longitude, along with census tract and block group information for each deliverable address in your database. You can also perform distance calculations based on a ZIP+4™.

The versatility is endless.



# DataMatch Enterprise Server Edition



**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 isn't an issue we would normally recommend a standard desktop machine which 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 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 and so 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. Deployment of data and temporary files is covered by a separate document on optimizing DataMatch Enterprise performance.

A SCSI interface will further increase the transfer rate from the disk and so is 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 very useful.

When fast performance is a priority we recommend using Solid State Drives (SSDs) as these are capable of very fast data access.

**Memory:** Although having more RAM than DataMatch Enterprise needs will not increase the speed of processing, having too little will decrease it dramatically. If running Windows XP we strongly recommend at least 2GB of free RAM (for a small/medium job), for larger jobs at least 4GB of free RAM is preferable.

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

## System Requirements

### Minimum

- Windows XP
- Pentium 4 (or equivalent)
- 2GB free RAM
- 300 MB Disk Space

### Optimum

- Windows 7
- Intel Core i7
- 12GB RAM
- SSD, or SCSI/SATA HDDs
- 100 GB Disk Space

## About Us

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.

## Success Stories

With over 2500 clients in 31 countries, Data Ladder has consistently gone above and beyond to ensure clients have the best technology for data cleansing and fuzzy record linkage.

*“Using DataMatch we were able to clean records from multiple systems and create a unified and deduplicated view of our customer. It’s not just the software, which works very well for us, but the focus and knowledge that Data Ladder brings to the table.”*

**J. Ciccone**  
**Data Quality Manager**  
**Hewlett Packard**



imagination at work



**HARVARD**  
UNIVERSITY



**ZURICH**

