**Entity Validation System (EVS)**

**A blue square with a white logo and a magnifying glass

AI-generated content may be incorrect.**

**User Guide**

**Table of Contents**

1. [Introduction](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#introduction)
2. [Installation](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#installation)
3. [Getting Started](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#getting-started)
4. [User Interface](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#user-interface)
5. [Workflow](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#workflow)
6. [Advanced Features](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#advanced-features)
7. [Troubleshooting](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#troubleshooting)
8. [FAQ](https://claude.ai/chat/b87312e5-4b79-41be-a048-ed48d6232edf#faq)

# Introduction

The Entity Validation System (EVS) is a comprehensive tool designed to standardize entity names and detect duplicates before importing data into enterprise systems. It performs multi-stage validation:

* Standardization of entity names
* Verification against naming conventions
* Duplicate detection
* Comparison with authoritative reference lists

# Installation

**System Requirements**

* Windows 10 or higher
* 4GB RAM minimum (8GB recommended)
* 500MB disk space

**Installation Steps**

1. Run the EntityValidationSystem\_Setup.exe installer
2. Follow the installation wizard
3. Launch from the Start menu or desktop shortcut

# Getting Started

**First Launch**

1. The application opens to the Setup tab
2. Default authoritative list (AuthoritativeEntityList.csv) is pre-configured
3. If needed, click "Browse..." to select a different authoritative list

**Preparing Input Files**

EVS works with CSV files containing entity data. Files should have at least one of these columns:

* name
* facility
* owner
* operator

Column names are case-insensitive.

# User Interface

**Main Window**

The EVS interface consists of four tabs representing the validation workflow:

**Setup Tab**

* **File Selection**: Select input CSV and authoritative list
* **Configuration**: Choose entity column and set matching threshold
* **Advanced Options**: Enable advanced matching algorithms

**Validation Tab**

* **Status Panel**: Shows validation progress and statistics
* **Violations Table**: Lists naming convention violations
* **Action Controls**: Accept or reject suggestions

**Duplicates Tab**

* **Internal Duplicates**: Find and resolve duplicates in your dataset
* **External Duplicates**: Compare with authoritative list

**Reports Tab**

* **Report Generation**: Create validation reports
* **Report List**: View and open generated reports

# Workflow

**Step 1: Setup**

1. Click "Browse..." to select your input CSV file
2. Verify or change the authoritative list location
3. Select the entity column (defaults to "name")
4. Adjust the matching threshold (85% recommended)
5. Click "Start Validation"

**Step 2: Naming Convention Validation**

1. Review validation results in the table
2. For each violation:
   * **Accept**: Use the suggested correction
   * **Reject**: Keep the original name
   * **Modify**: Edit the suggestion
3. Use "Accept All" or "Reject All" for batch processing
4. Click "Apply Changes" to save your decisions
5. Click "Proceed to Duplicate Detection"

**Step 3: Duplicate Detection**

1. Review Internal Duplicates:
   * Entities that appear to be duplicates within your dataset
   * Choose "Keep Best", "Merge", or "Custom" for each group
   * Click "Resolve Selected Group" or "Resolve All Groups"
2. Review External Duplicates:
   * Entities that match entries in the authoritative list
   * Choose "Accept Match", "Reject Match", or "New Entity"
   * Click "Accept All Matches" or "Reject All Matches" for batch processing
3. Click "Apply Changes" to save decisions
4. Click "Continue to Reports"

**Step 4: Report Generation**

1. Review report options
2. Enable "Update original file with validated entities" if desired
3. Click "Generate Reports"
4. View reports in the list
5. Click "Open Selected Report" or "Open Reports Folder"

# Advanced Features

**Matching Threshold Configuration**

* **Higher threshold (90%+)**: Fewer false positives, may miss some duplicates
* **Medium threshold (80-90%)**: Balanced approach
* **Lower threshold (70-80%)**: Finds more potential matches, more false positives

**Advanced Matching**

When enabled, uses additional algorithms:

* Levenshtein Distance
* Token-Based Matching
* Phonetic Matching
* Machine Learning-Based Matching (requires more processing time)

**Naming Convention Rules**

EVS enforces these standards:

1. Full names instead of acronyms
2. "and" instead of "&"
3. No apostrophes
4. No special characters
5. No business designations (LLC, Inc., etc.)
6. Expanded abbreviations (Co -> Company)
7. Government entity formatting (State of X, County of Y (ST))

# Troubleshooting

**Common Issues**

**Application Won't Start**

* Verify you have administrator rights
* Check if antivirus is blocking execution
* Reinstall the application

**File Loading Errors**

* Ensure CSV is properly formatted
* Check that the file isn't open in another application
* Verify the CSV contains required columns

**Slow Performance**

* Reduce dataset size
* Disable advanced matching for large datasets
* Increase system memory

**Output File Issues**

* Check write permissions to the reports directory
* Close any open report files before regenerating

# FAQ

**Q: How many entities can EVS process?** A: EVS can handle datasets with up to 100,000 entities, though performance depends on hardware.

**Q: Can I customize naming convention rules?** A: The current version uses fixed rules. Contact your system administrator for customization.

**Q: Does EVS modify my original files?** A: No, EVS creates new files unless you specifically enable the "Update original file" option.

**Q: How accurate is duplicate detection?** A: Accuracy depends on the matching threshold and data quality. Start with the default 85% and adjust as needed.

**Q: Where are reports saved?** A: Reports are saved in the "reports" folder within the application directory.