# Overview

The Deduplication and Official Linkage Engine (DOLE) is a Windows desktop application designed to detect duplicates, link user records to master databases, and flag government officials in beneficiary lists. This comprehensive tool provides high-accuracy name matching with nickname equivalence, phonetics, and fuzzy ratios to ensure data integrity in government programs.

# System Requirements

* **Operating System:** 64-bit Windows 8.1, 10, or 11 (or Windows Server 2012 R2+)
* **Installation:** No installation required - run directly from any folder or USB drive
* **Optional:** Microsoft Excel (desktop) for PDF export functionality
* **Internet:** Required for initial setup and periodic updates (offline mode available)

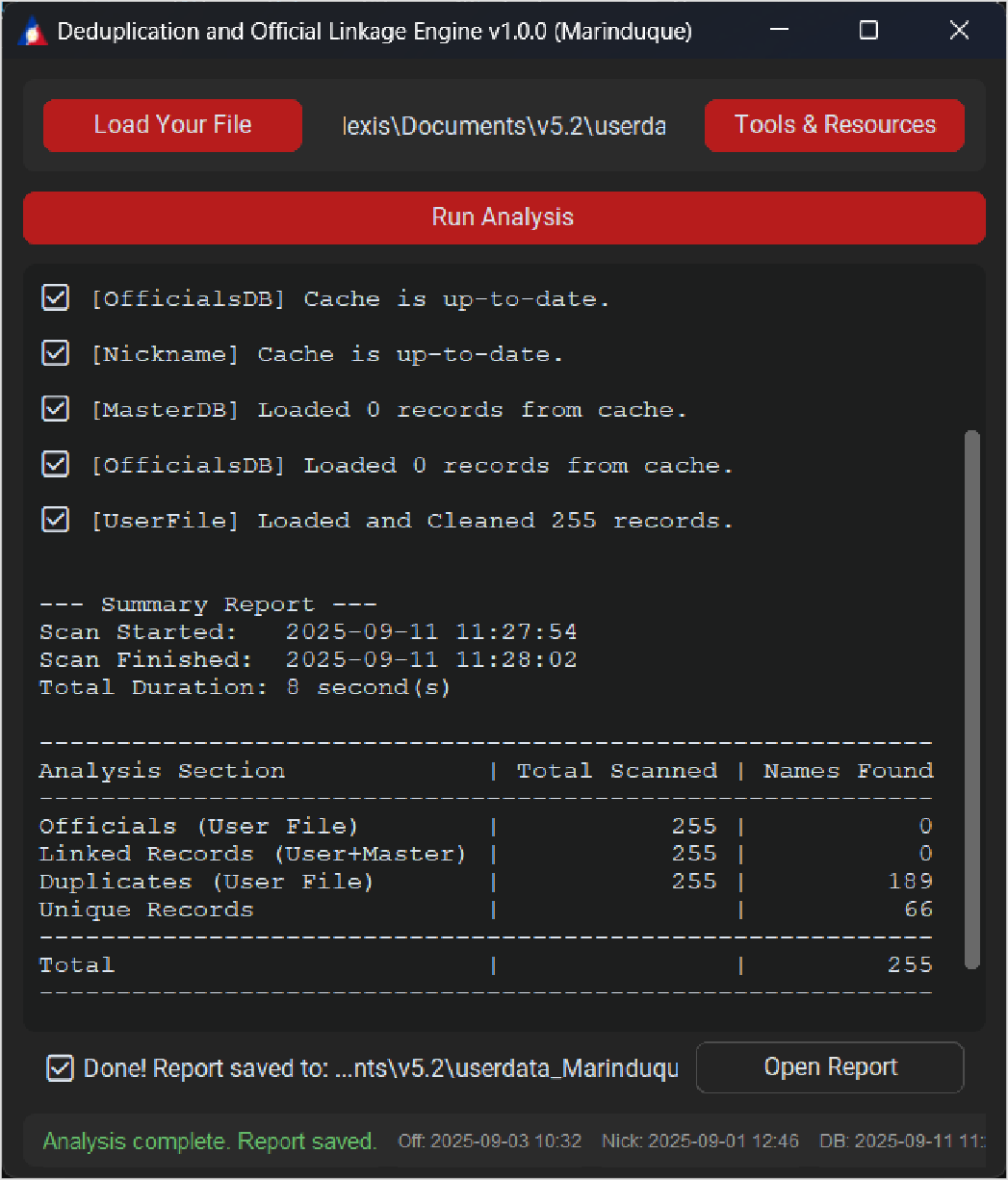
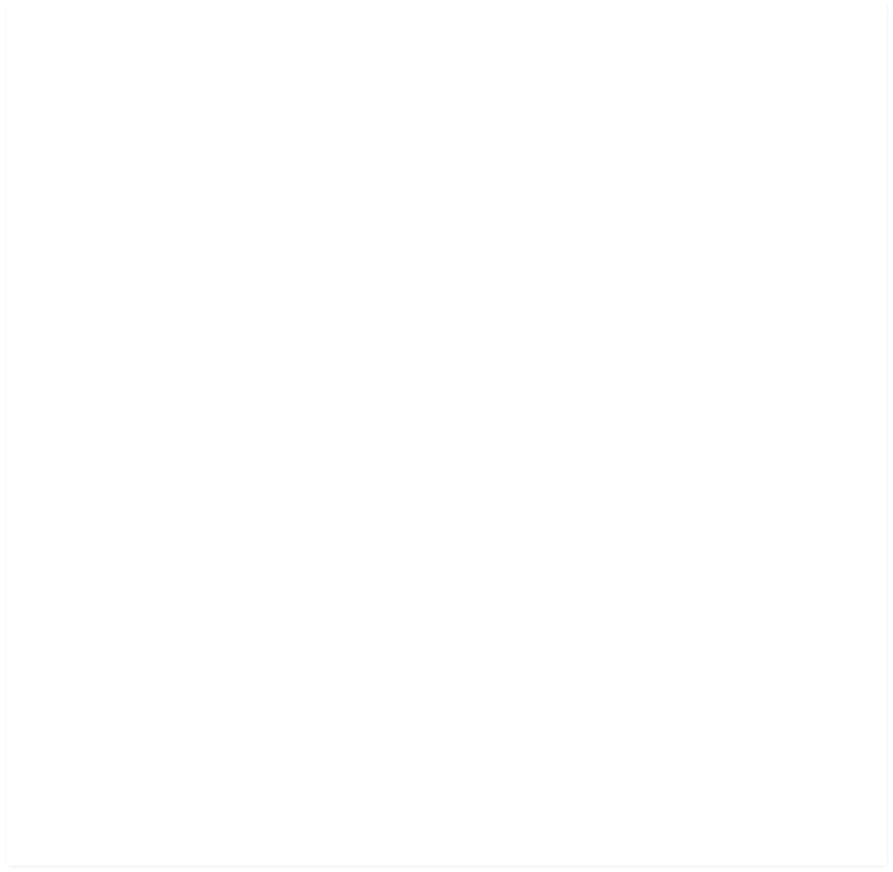
# Getting Started

## Launching the Application

Run **DOLE\_v{{VERSION}}\_{{PROVINCE}}.exe** to launch the main application.

## Main Interface Overview

The main application window provides all the essential tools for data analysis:



*Figure 1.1*

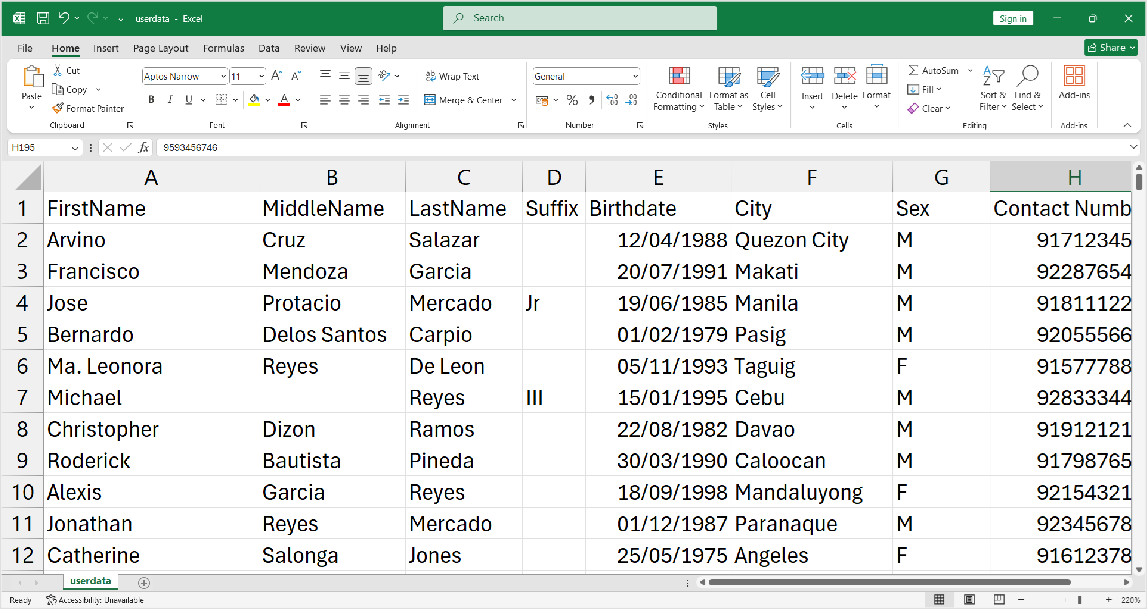
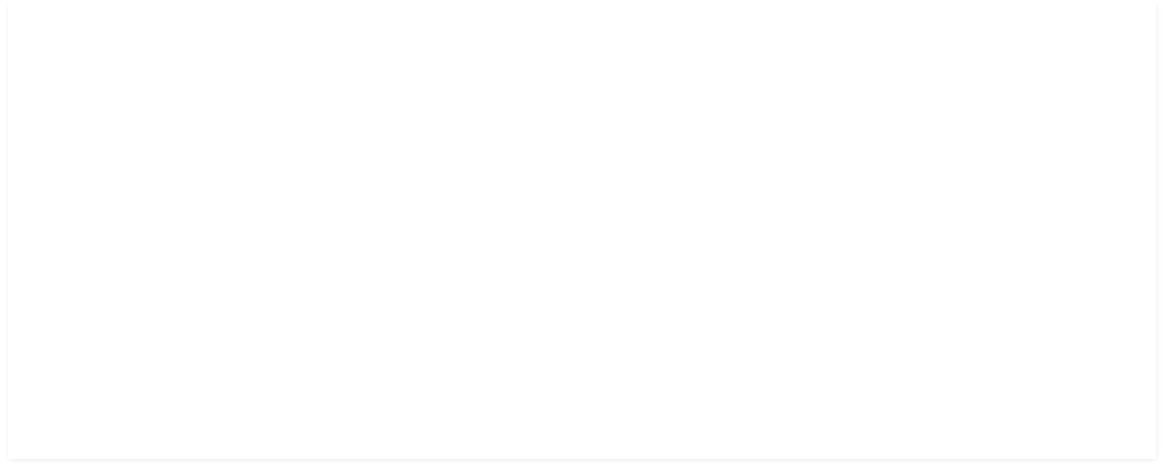
Interface Components:

1. **Load Button** - Select your data file for analysis
2. **Tools & Resources** - Access configuration and utility options
3. **Run Analysis** - Start the deduplication process
4. **Progress Bar** - Shows real-time analysis progress
5. **Open Report Button** - Quick access to generated reports
6. **GUI Logs** - View detailed processing information and status updates
7. **Footer Information** - Application status, version details, and system information

# Step-by-Step Usage Guide

# Step 1: Preparing Your Data

Before starting analysis, ensure your data follows the correct format:



*Figure 1.2*

Required Columns:



1. **First Name** - Individual's given name
2. **Middle Name** - Middle name or initial
3. **Last Name** - Family name or surname
4. **Birthdate** - Date of birth (consistent format recommended)
5. **City** - City or municipality of residence
6. **Sex** - Gender (M/F or Male/Female)
7. **Contact Number** - Phone number (preserved as-is for follow-up)

## Step 2: Loading Your Data File

1. Click the Load button in the main interface
2. Browse and select your input file

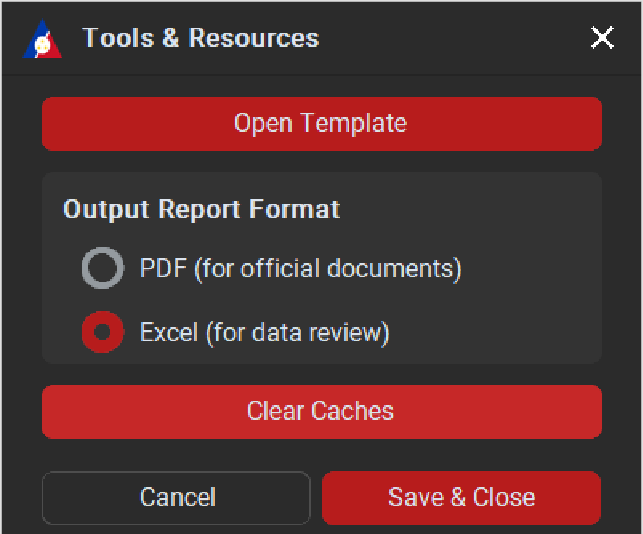
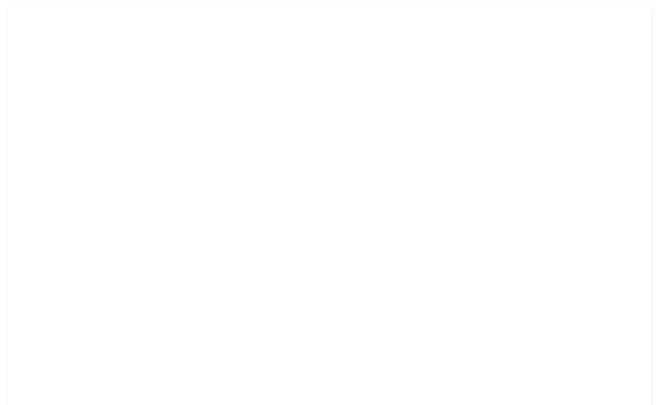
### Supported formats:

* + CSV files (.csv)
  + Excel files (.xlsx)
  + Text files (.txt)

The application will validate your file structure and display any formatting issues in the GUI logs.

## Step 3: Configuring Application Settings

Click Tools & Resources to access configuration options:



*Figure 1.3*

Available Configuration Options:

1. **Open Template** - View the standardized data format template
2. **Output Report Format** - Choose between PDF or Excel output formats
3. **Clear Cache** - Remove stored reference files and cached data
4. **Cancel/Save Buttons** - Manage your configuration changes

### Using the Data Template

*Figure 1.4*

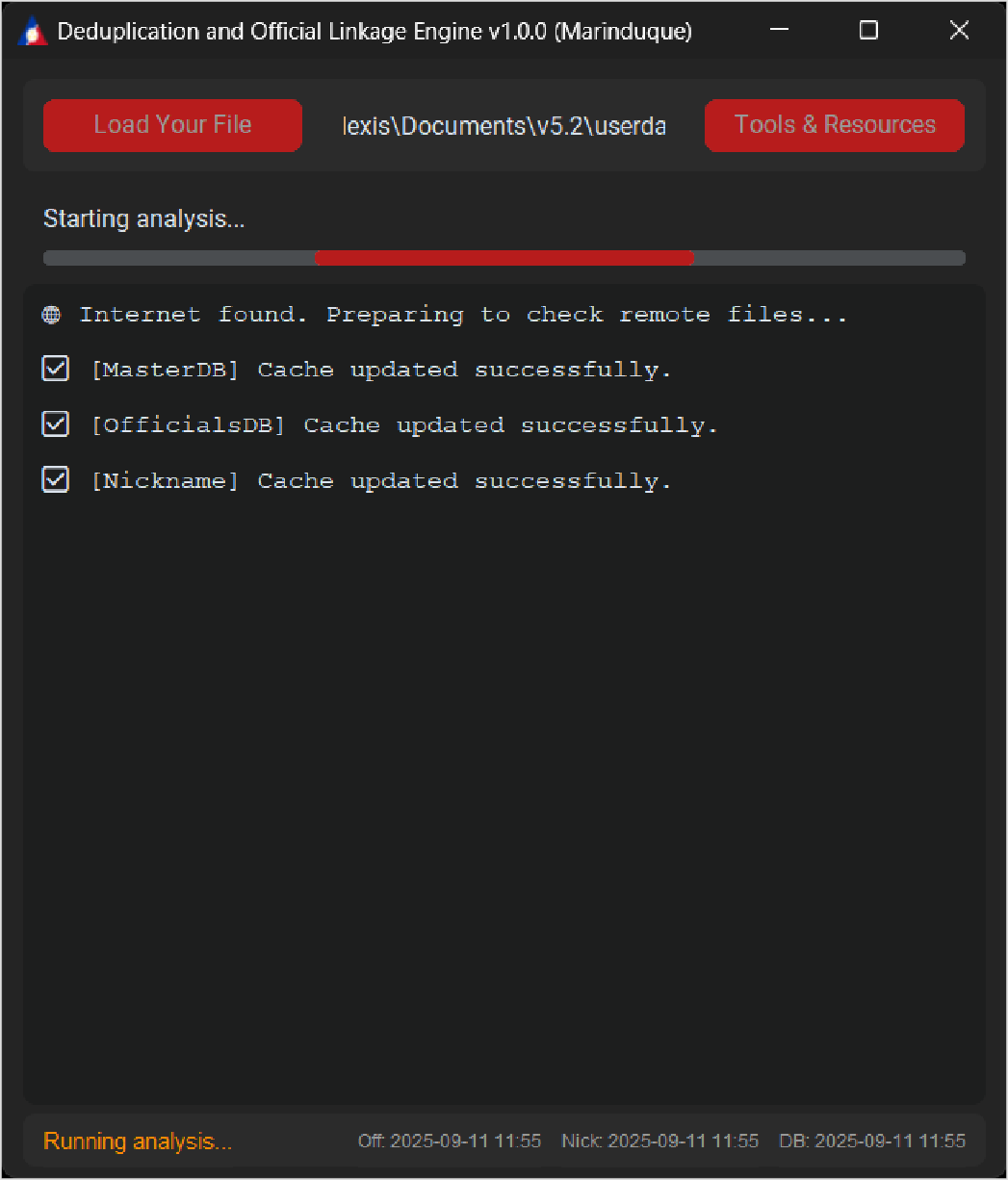
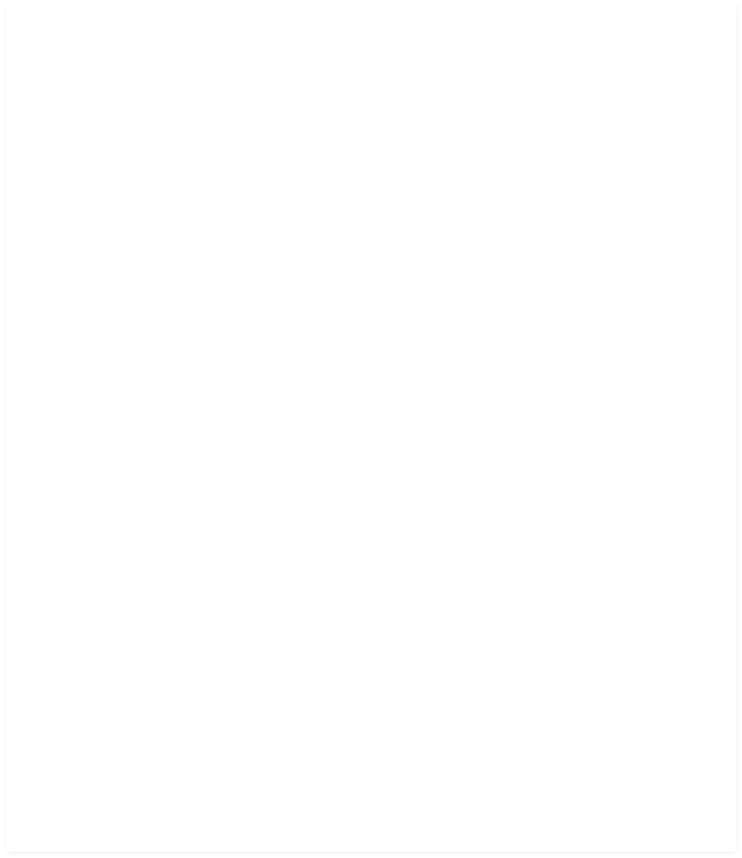
The template provides the exact column structure and formatting requirements for optimal results. Use this as a reference when preparing your data files.

**Step 4: Running the Analysis**

The application operates in different modes based on internet connectivity:

### Online Mode (Recommended)

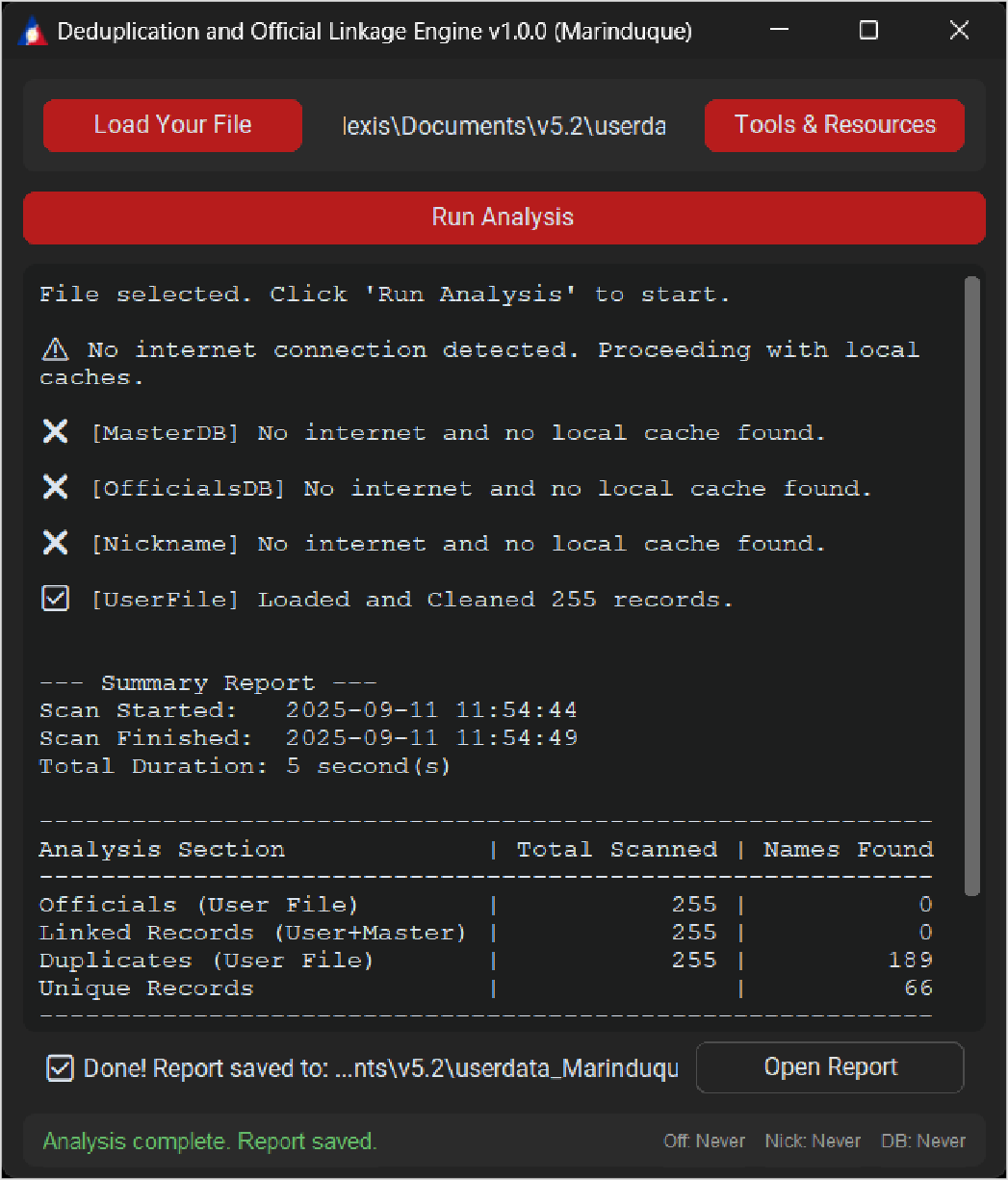
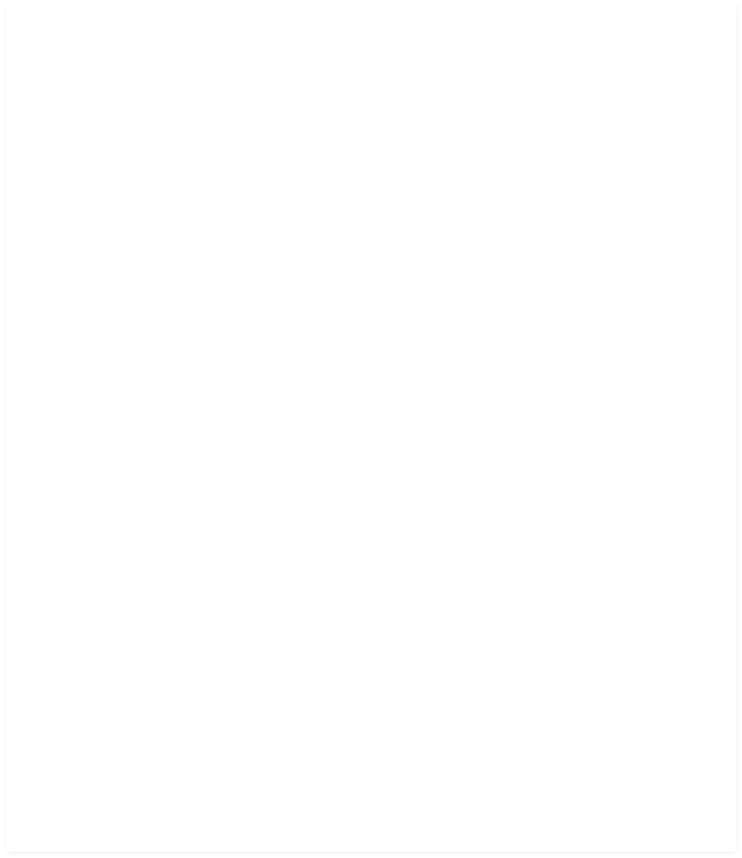
*Figure 1.5*



**When internet is available:** - Downloads latest reference databases - Enables full matching capabilities - Provides most accurate results - Updates cached files automatically

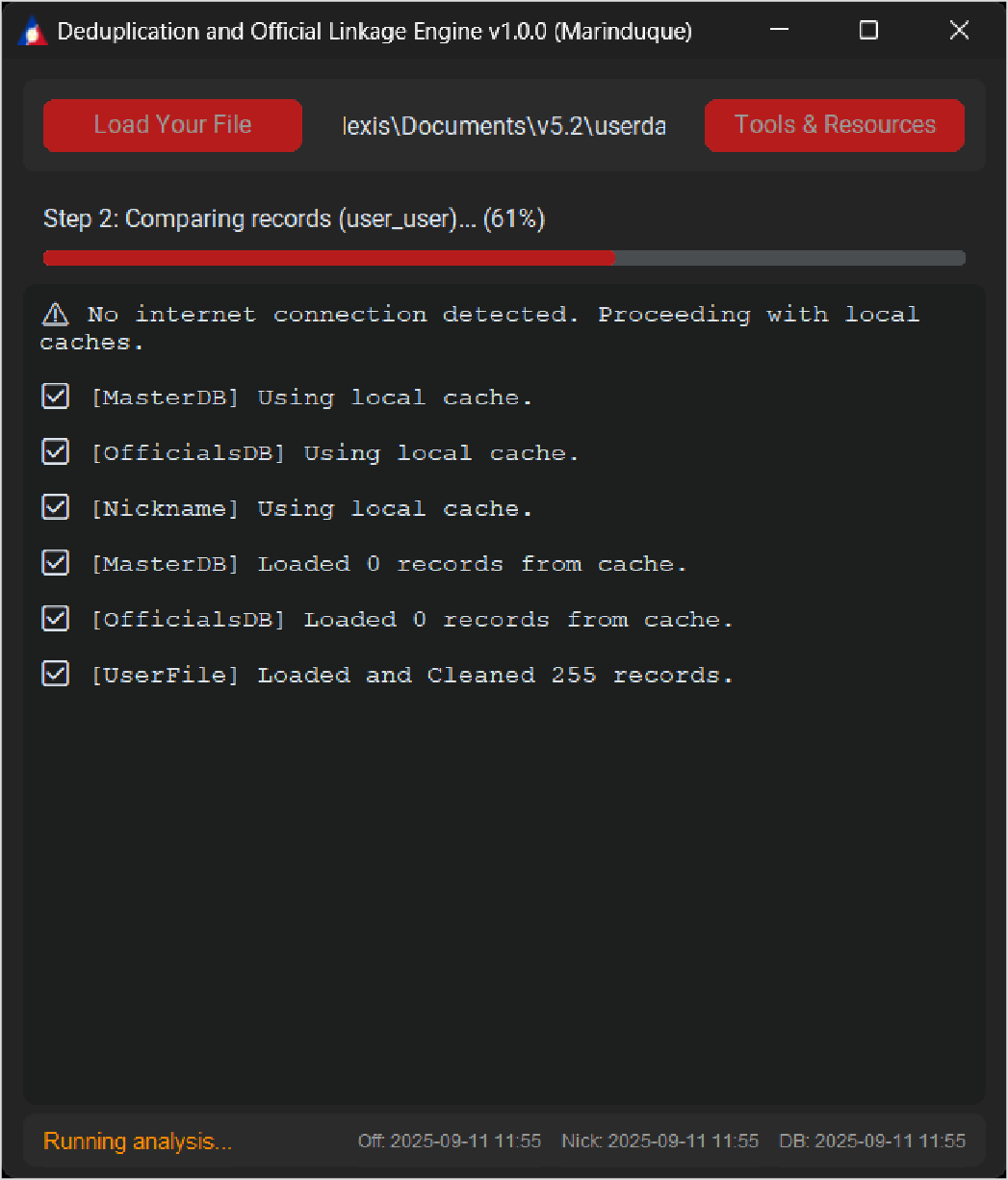
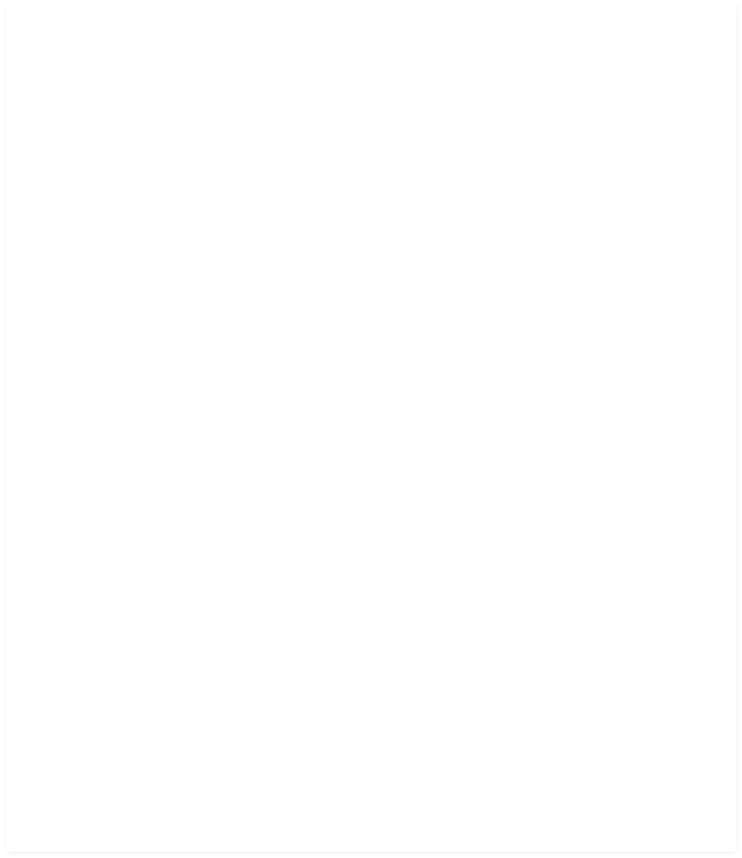
### Offline Mode - First Run

*Figure 1.6*



**Without internet connection:** - Limited to duplicate detection only - No external database matching - Reduced functionality until cache is established  
**Offline Mode - With Cache**

*Figure 1.7*



**After initial online setup:**

* Full functionality available offline
* Uses previously cached reference files
* Automatic update check when internet resumes
* Maintains data integrity and accuracy

## Step 5: Monitoring Analysis Progress

### Understanding GUI Logs

The GUI Logs section provides comprehensive real-time information:

Log Information Includes:

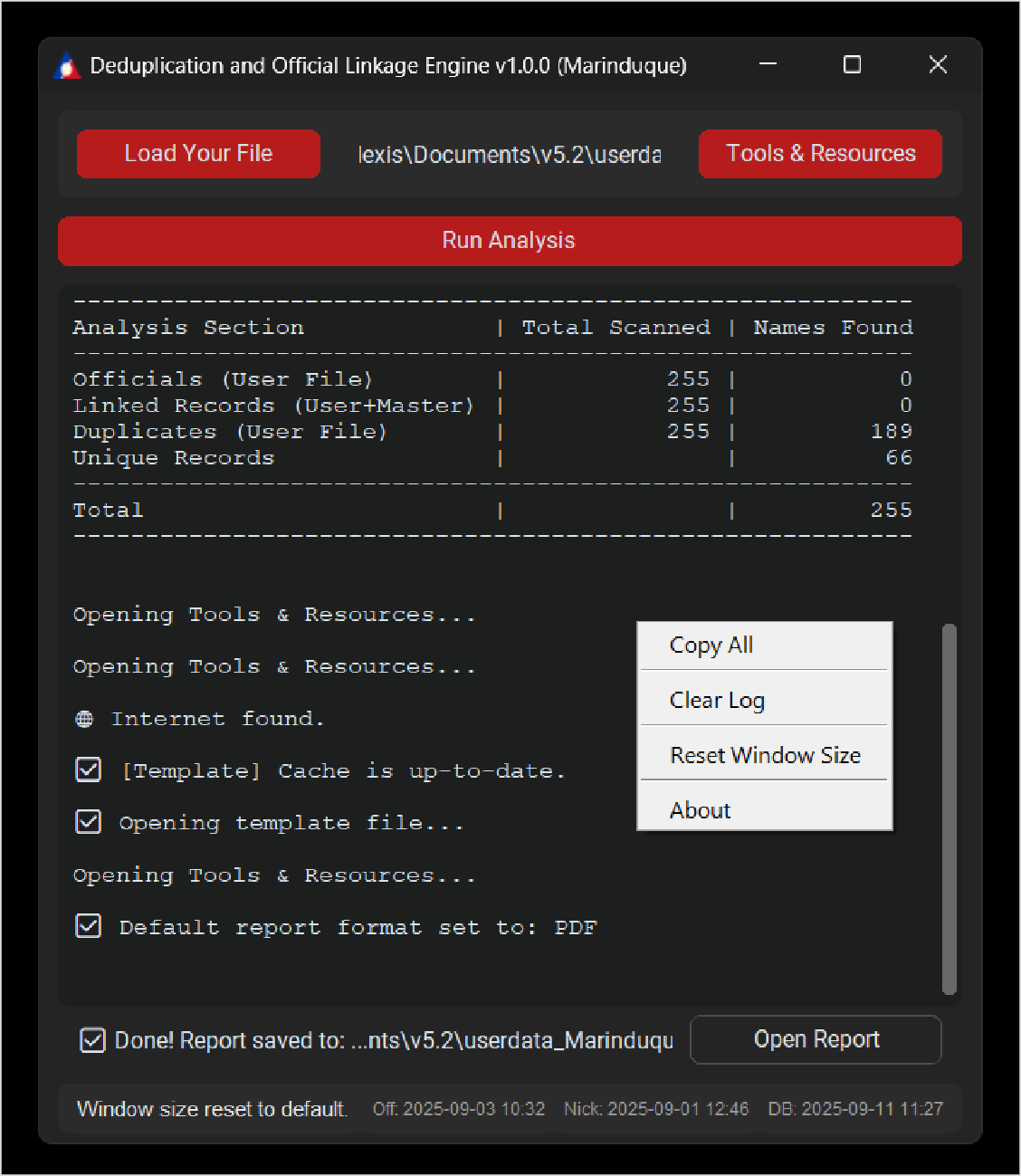
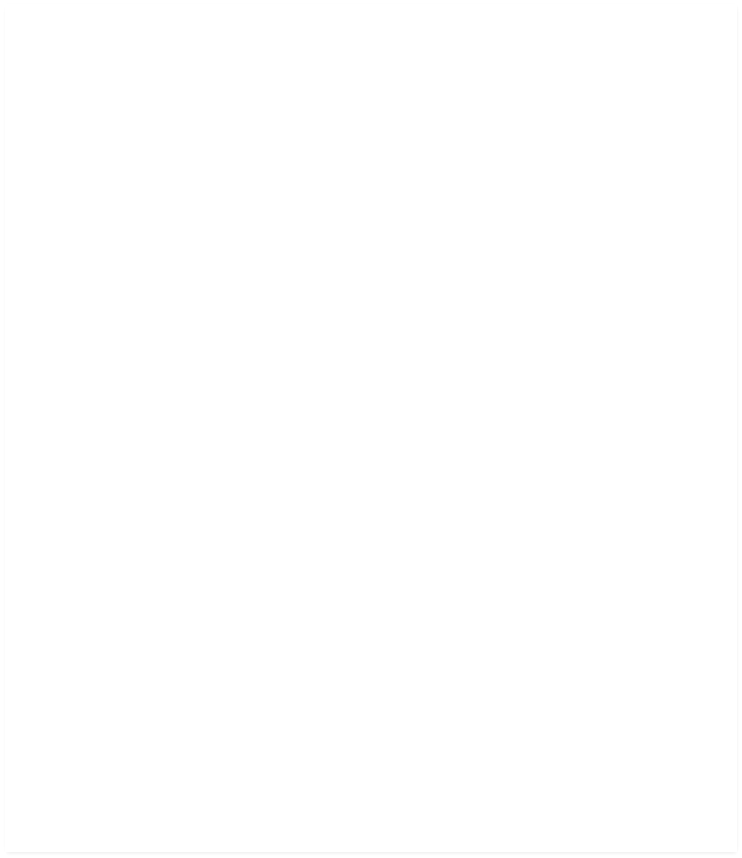
* **File Loading Status** - Confirmation of successful data import
* **Analysis Progress** - Step-by-step processing updates
* **Match Detection Results** - Real-time matching statistics
* **Error Messages** - Detailed error descriptions and solutions
* **Completion Status** - Final analysis summary and results location

### Summary Report in GUI Logs

After analysis completion, the GUI logs display a comprehensive summary report showing:

* **Scan Duration** - Total time taken for analysis
* **Analysis Sections** - Breakdown of different matching categories:
* **Officials (User File)** - Government officials found in your data
* **Linked Records (User+Master)** - Records matched to master database
* **Duplicates (User File)** - Duplicate entries within your dataset
* **Unique Records** - Records with no matches found

### GUI Logs Context Menu *Figure 1.8*

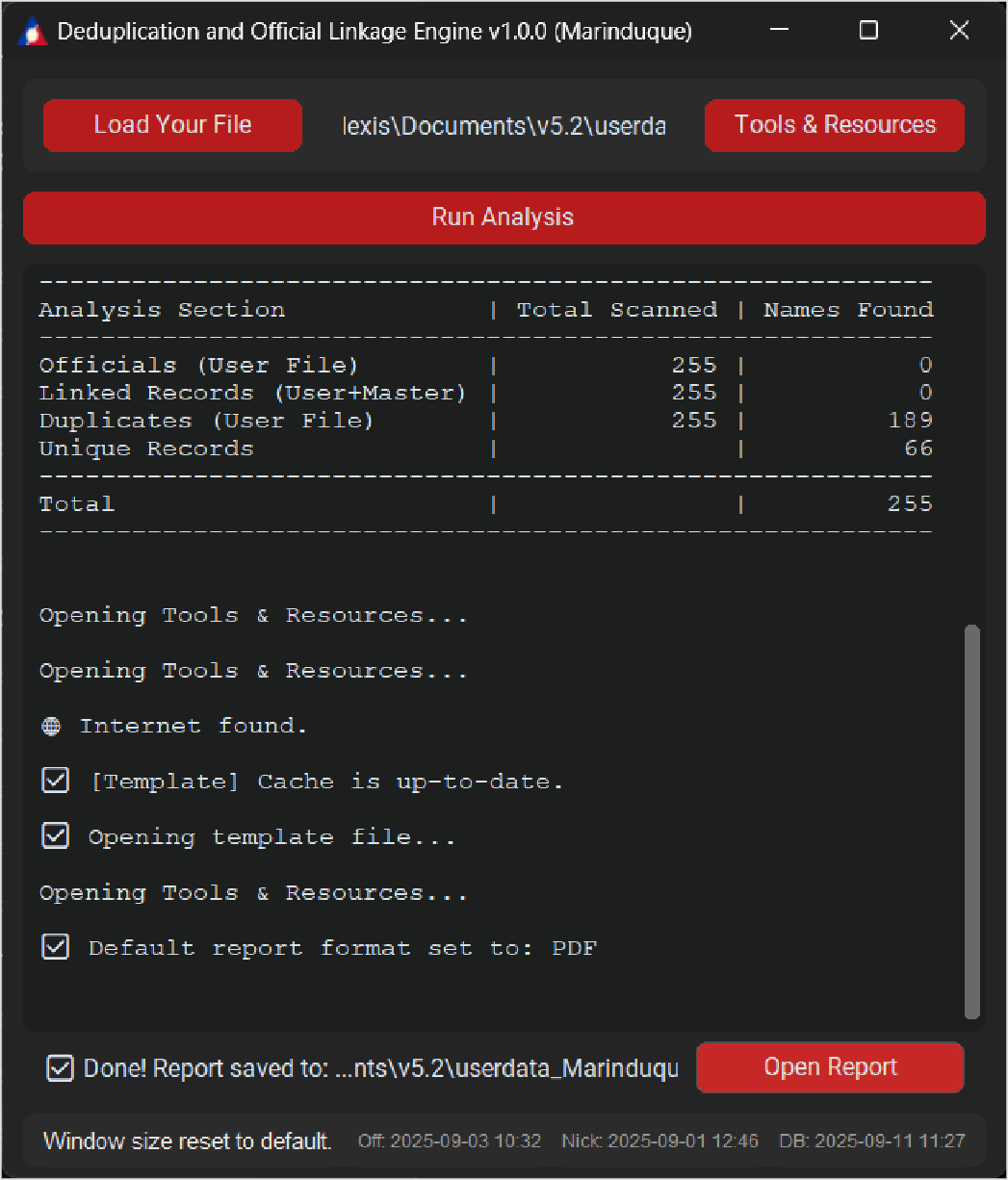
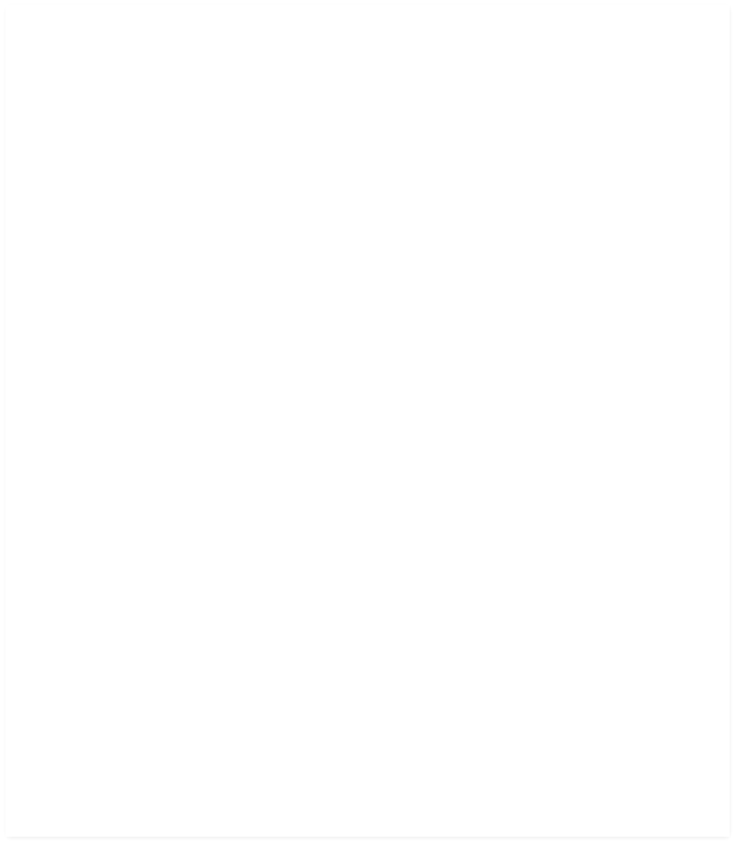


Right-click on the logs area to access:

* **Copy All** - Copy complete log content to clipboard for sharing or documentation
* **Clear Log** - Remove all current log entries for a fresh start
* **Reset Window Size** - Restore application window to default dimensions
* **About** - View detailed application information and credits

## Step 6: Accessing and Understanding Reports

### Opening Reports After analysis completion, use the Open Report Button for immediate access:



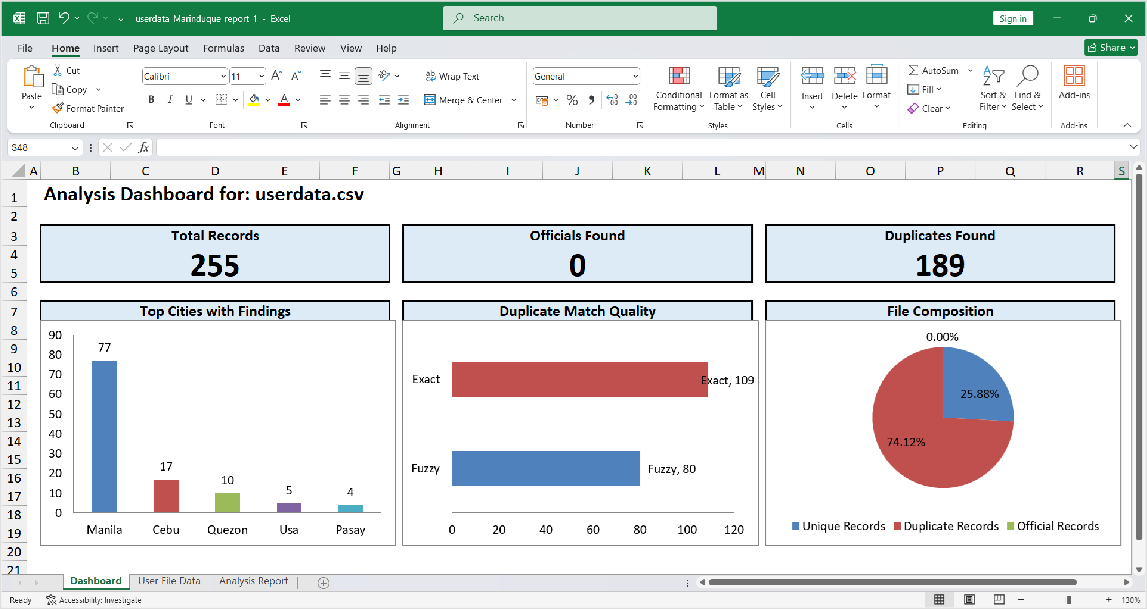
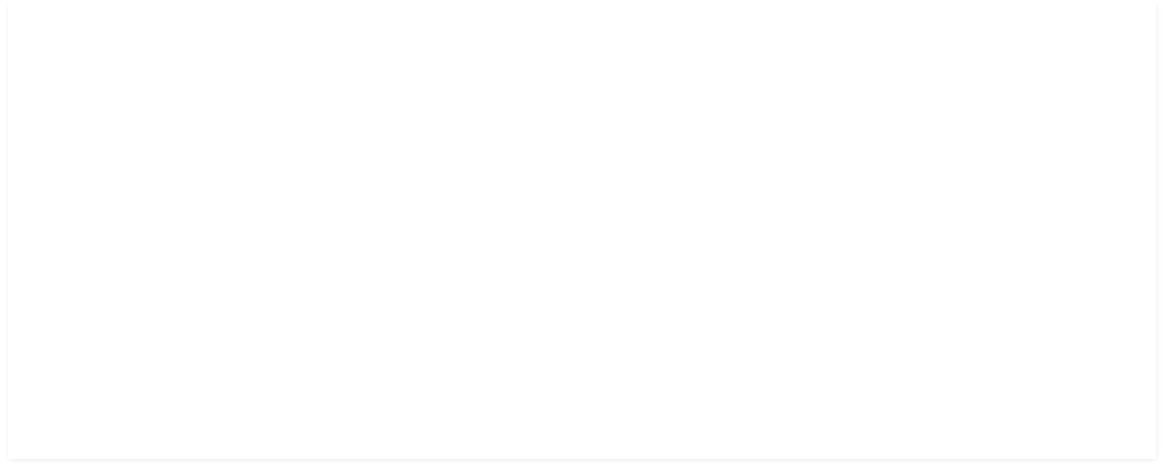
*Figure 1.9*  
  
  
  
  
  
  
  
Report Location: Reports are automatically saved in the same directory as your input file using the naming convention:

<input\_filename>\_<Province>\_report\_<N>.xlsx

# Comprehensive Report Analysis

## Excel Report Structure

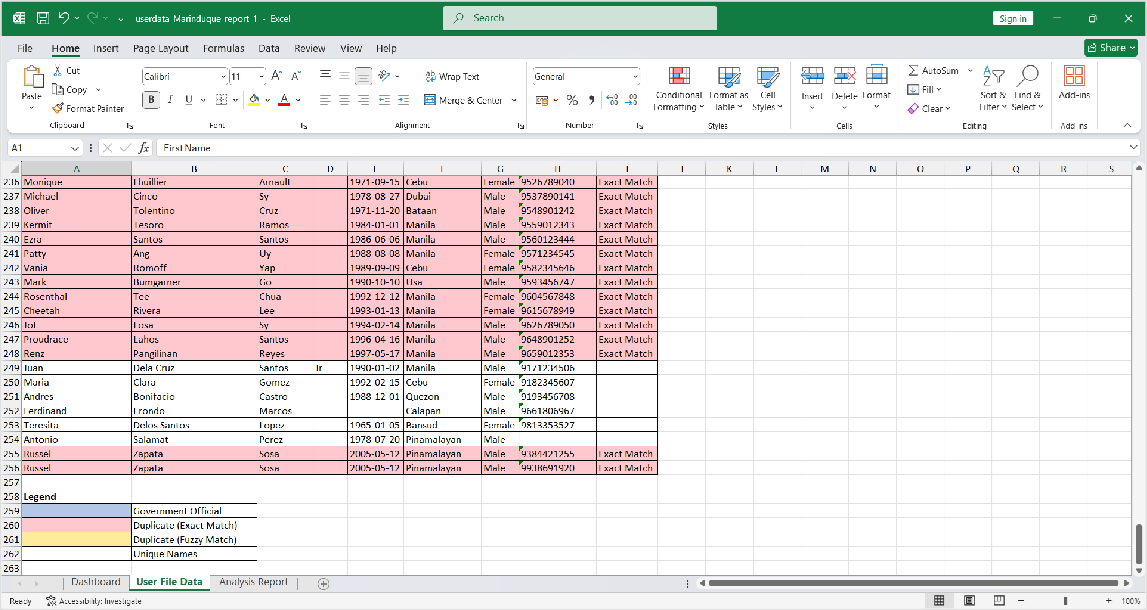
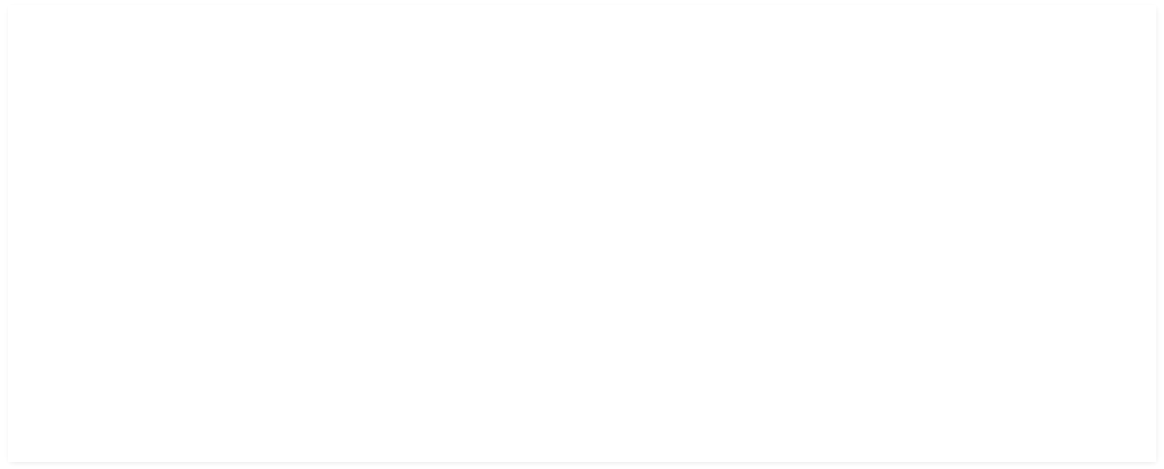
### Dashboard Sheet - Key Performance Indicators



*Figure 2.0*

Dashboard Metrics Include: **Total Records Processed** - Complete count of analyzed entries, **Officials Found** - Number of government officials identified, **Duplicates Found** - Count of duplicate records detected, **Top Cities with Findings** - Geographic distribution of matches, **Duplicate Match Quality** - Breakdown of exact vs fuzzy matches, **File Composition** - Visual representation of data categories

### User File Data Sheet - Detailed Results

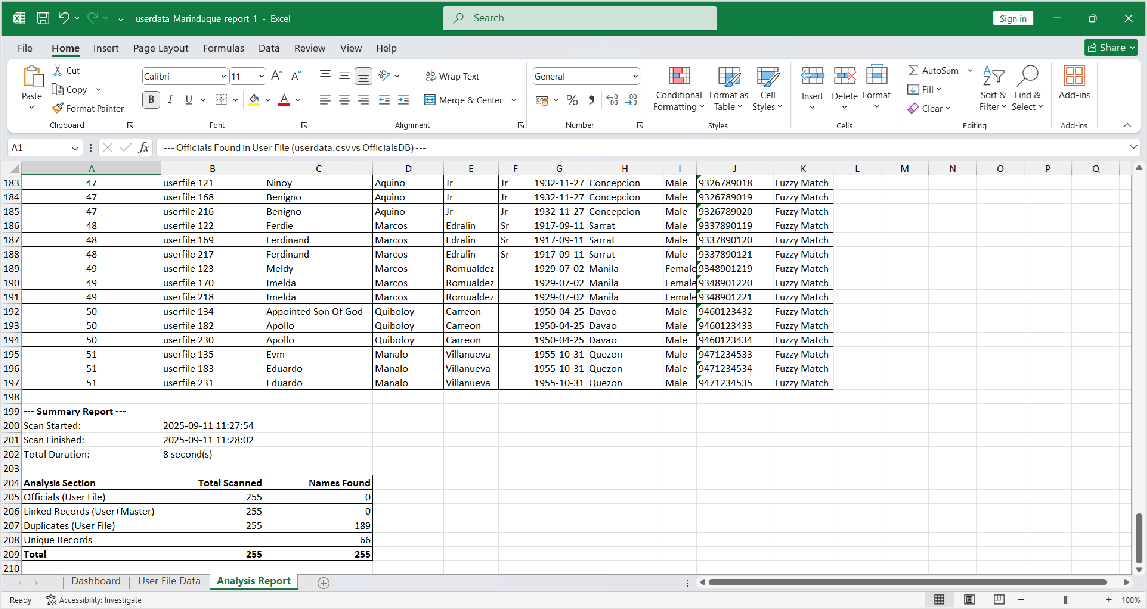
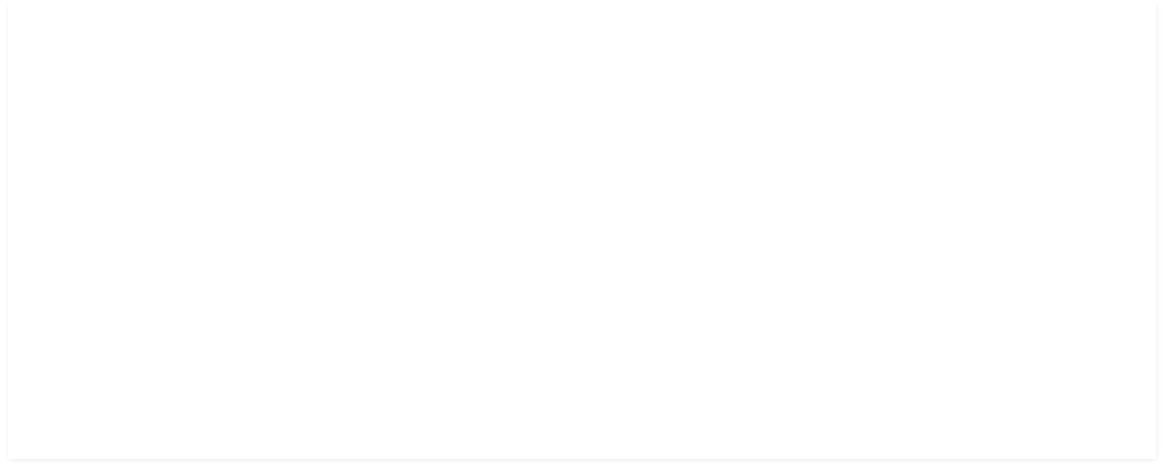


*Figure 2.1*

Color-Coded Legend System: - □ Blue (Government Official): Confirmed government officials - □ Red (Duplicate Exact Match): Exact duplicate matches - □ Yellow (Duplicate Fuzzy Match): Fuzzy duplicate matches - □ White (Unique Names): No matches found

### Analysis Report Sheet - Comprehensive Findings

*Figure 2.2*

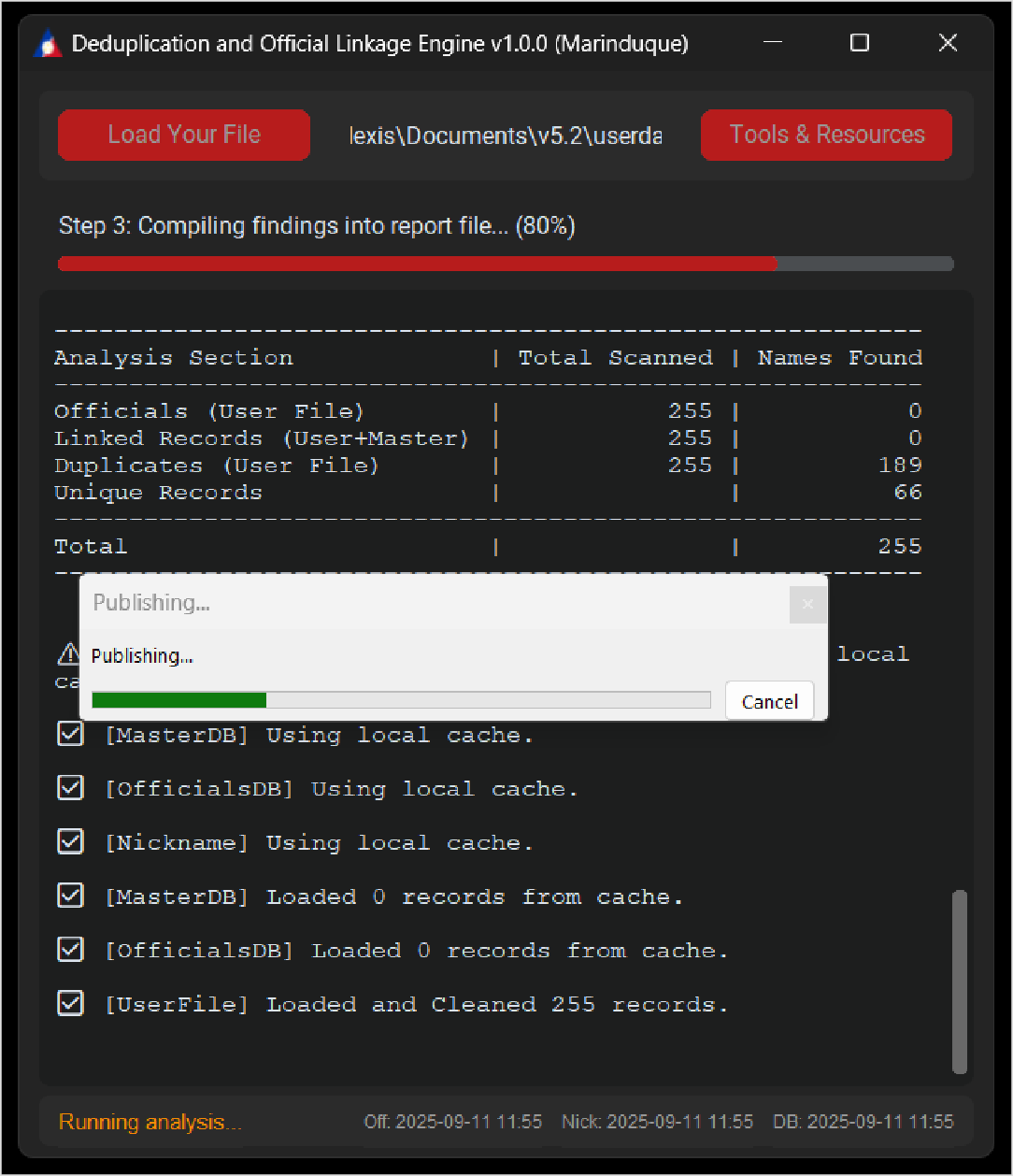
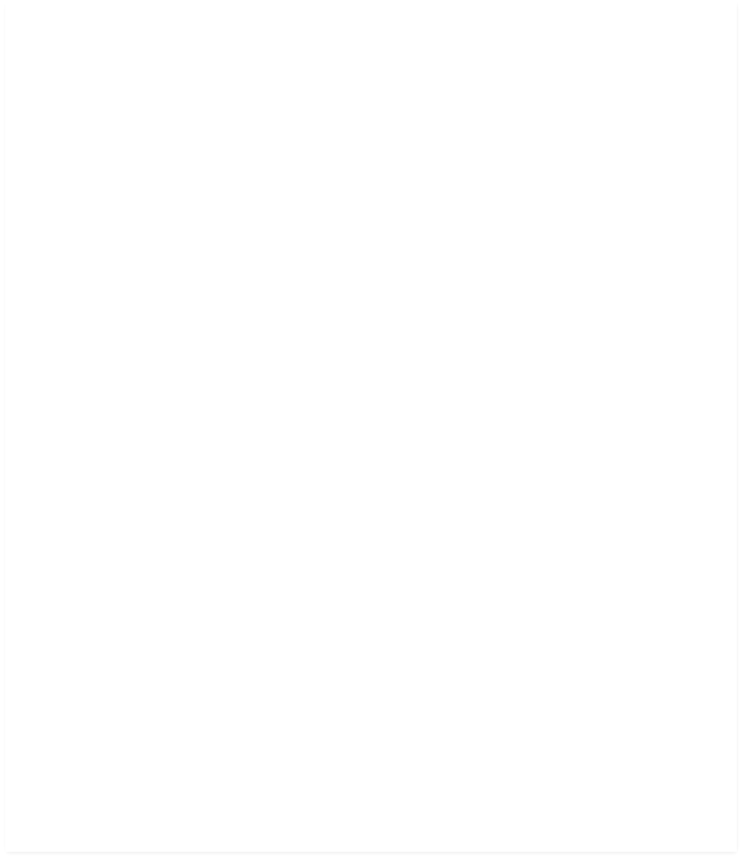


Detailed Analysis Components: **Match Classifications** - Categorized matching results, **Similarity Scores** - Numerical confidence ratings, **Connection Details** - How records are

linked, **Summary Statistics** - Final analysis totals

## PDF Report Format

For organizations with Microsoft Excel licenses:



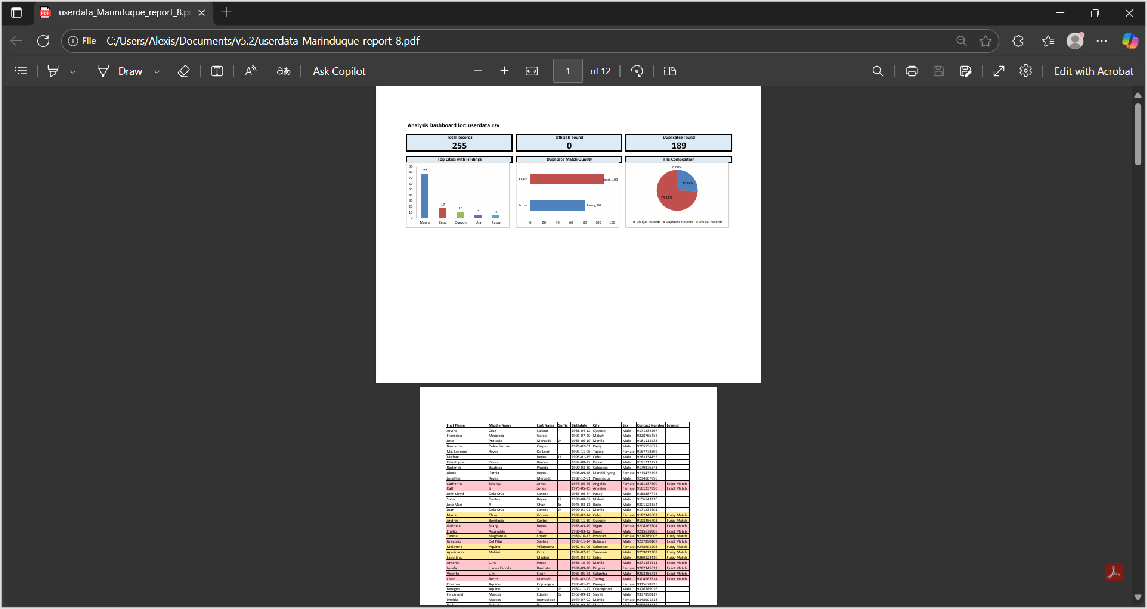
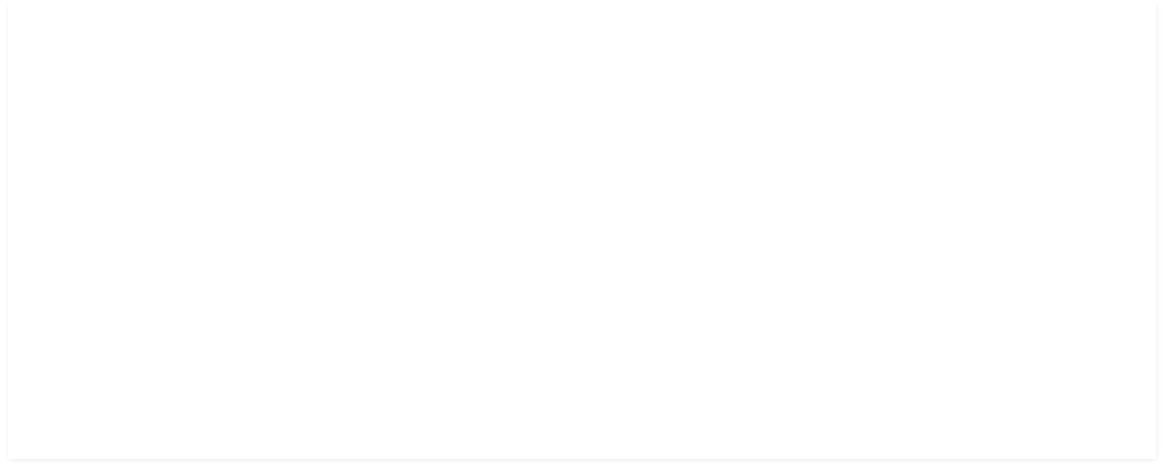
*Figure 2.3*

**PDF Report Structure**

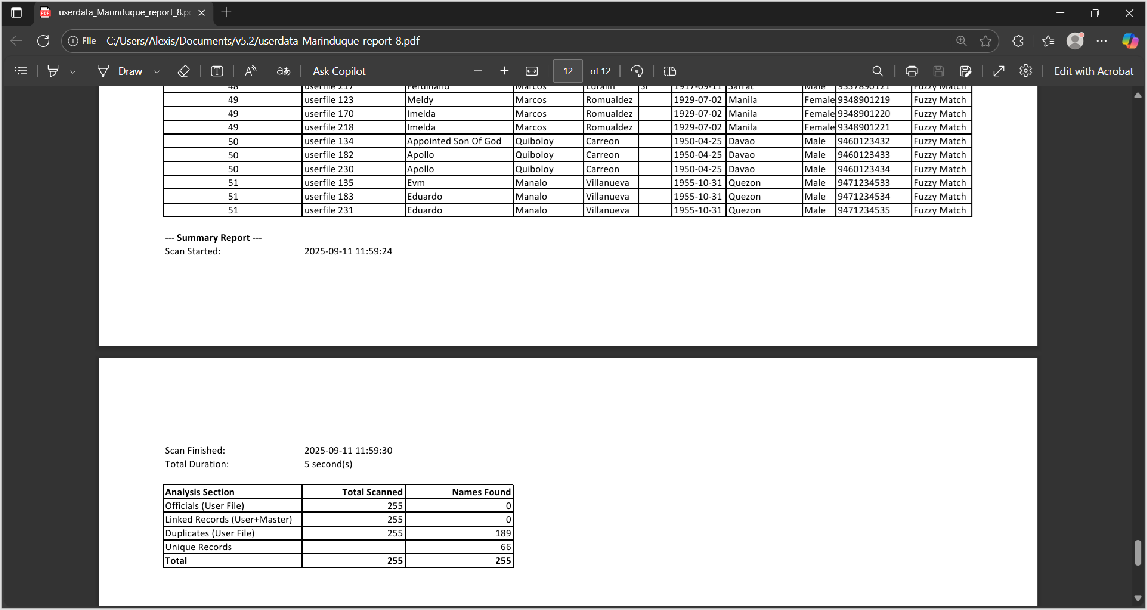
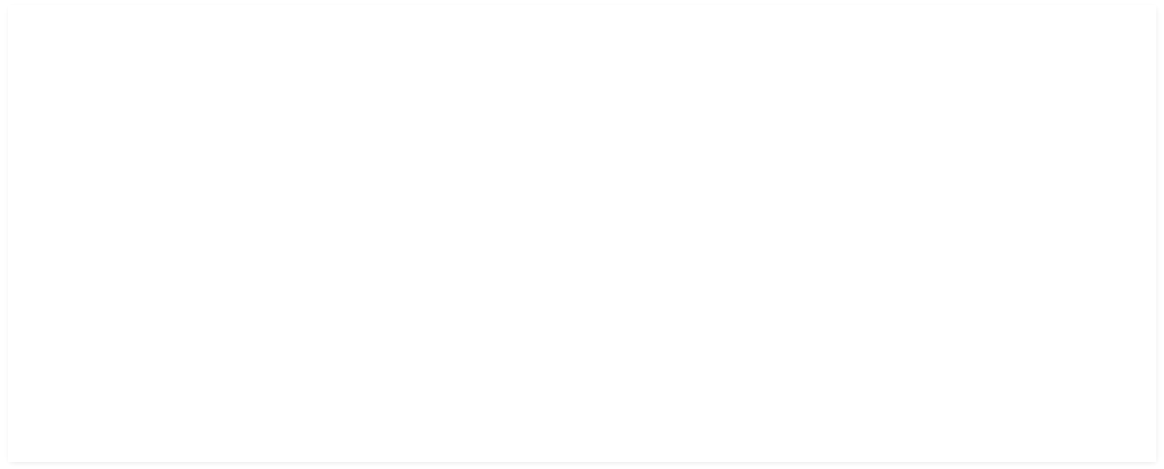


Page 1: Dashboard and Summary Data

*Figure 2.4*



Page 2: Detailed Analysis Results



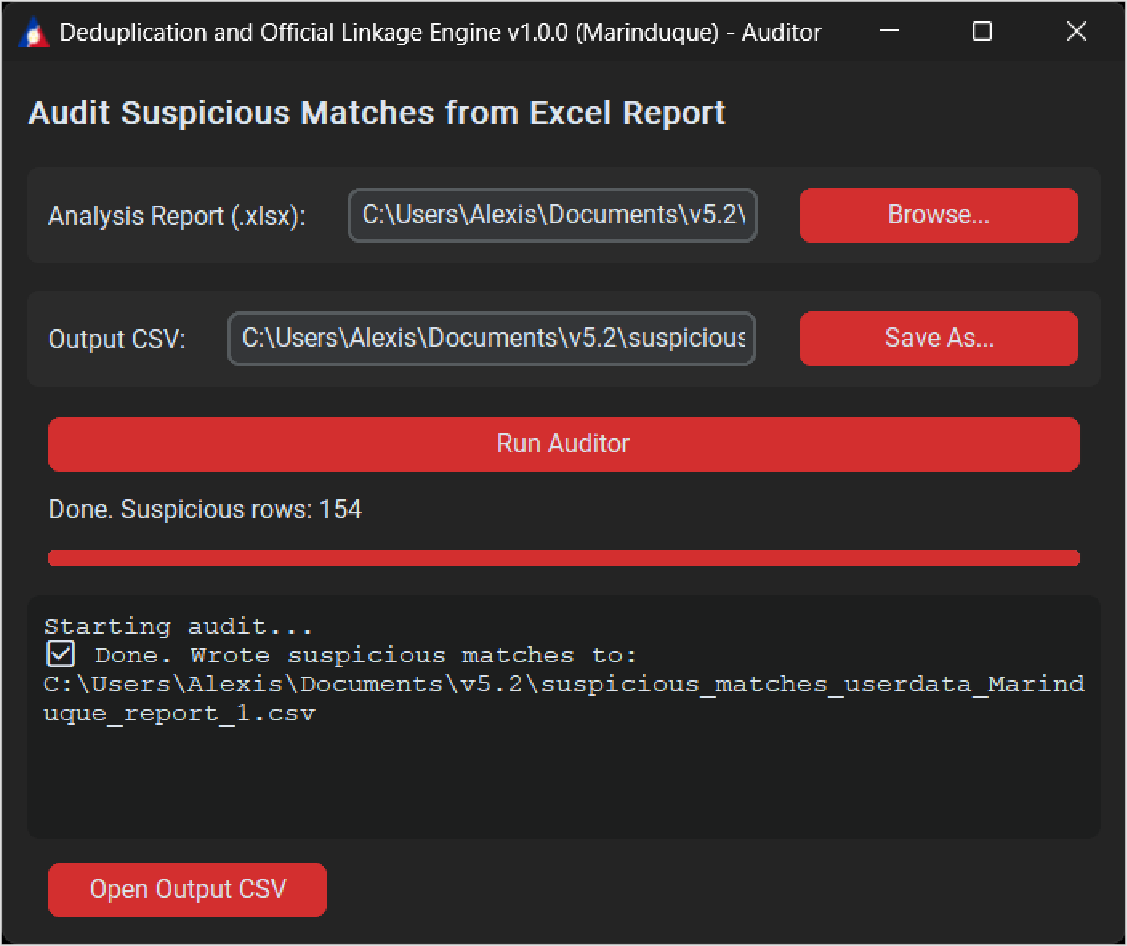
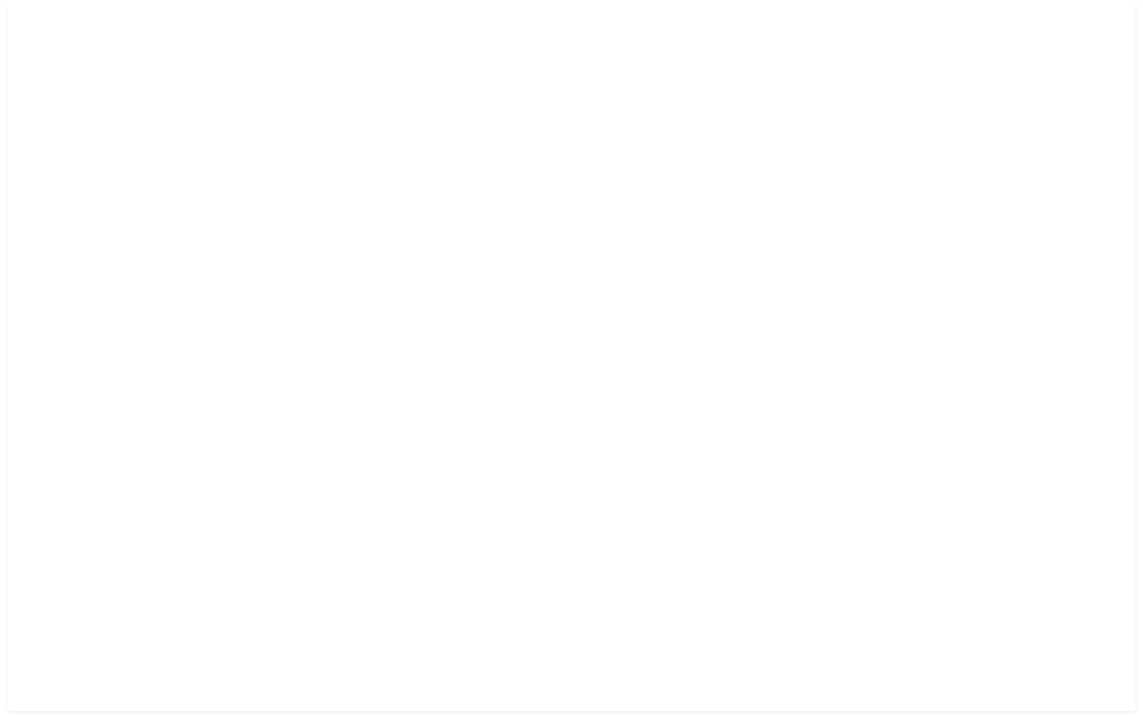
*Figure 2.5*

**PDF Benefits:** - Professional presentation format - Easy sharing and distribution - Preserved formatting across systems - Suitable for official documentation

**Advanced Features: Auditor Tool**

## Launching the Auditor Application

Run **DOLE\_v{{VERSION}}\_{{PROVINCE}}\_Auditor.exe** for advanced analysis review:



*Figure 2.6*



## Auditor Interface Components

Main Interface Elements:

1. **Analysis Report Browse Button** - Select Excel report from main application
2. **Output CSV Save As Button** - Specify location for suspicious matches’ file
3. **Run Auditor Button** - Execute the auditing process
4. **Progress Bar** - Real-time auditing progress indicator
5. **GUI Logs** - Detailed auditing process information
6. **Open Output CSV** - Direct access to generated suspicious matches report

## Auditor Workflow

### Step-by-Step Process:

1. Select Input Report
2. Click Browse to choose an Analysis Report (.xlsx) from the main application
3. Verify the report contains the required analysis data

### Configure Output

1. Click Save As to specify output location and filename
2. Choose an accessible location for easy review

### Run Analysis

1. Click Run Auditor to start the suspicious matches detection
2. Monitor progress through the progress bar and logs

### Review Results

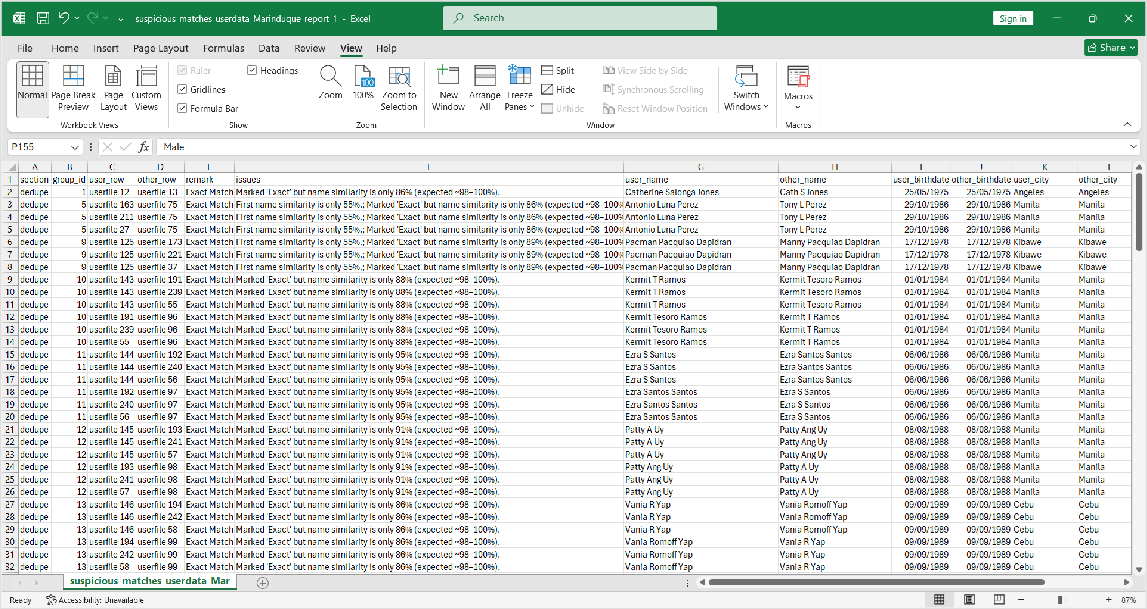
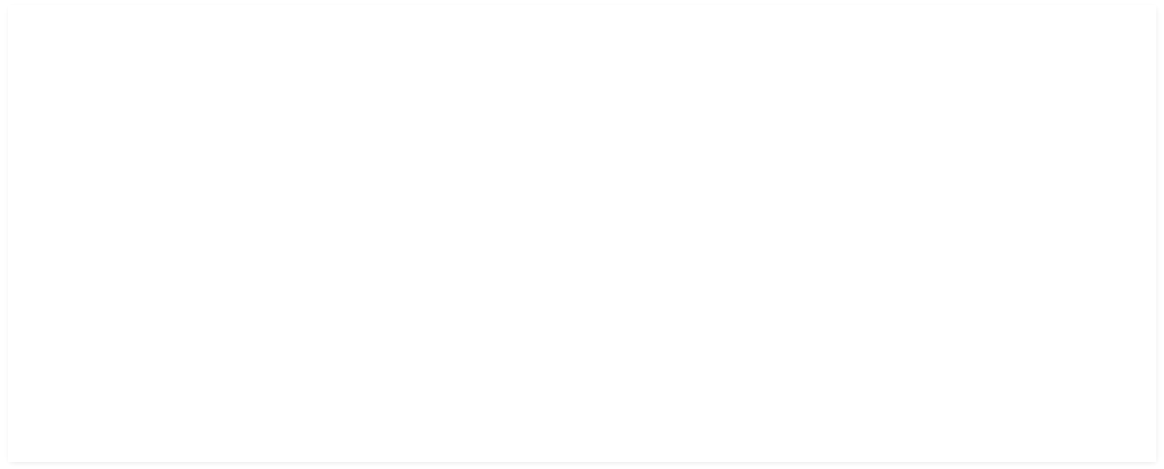
1. Click Open Output CSV when analysis completes
2. Review flagged records for manual verification

## What the Auditor Flags

The auditor automatically identifies records with potential issues: **Birthdate/Sex/Suffix Mismatches**

- Inconsistent personal information, **Weak Name Similarity** - Low confidence matches with middle initial conflicts, **Geographic Inconsistencies** - City mismatches when only names match, **Classification Anomalies** - "Exact" matches with unexpectedly low similarity scores

## Suspicious Matches Review



*Figure 2.7*

### Output File Details:

suspicious\_matches\_<report\_base>.csv

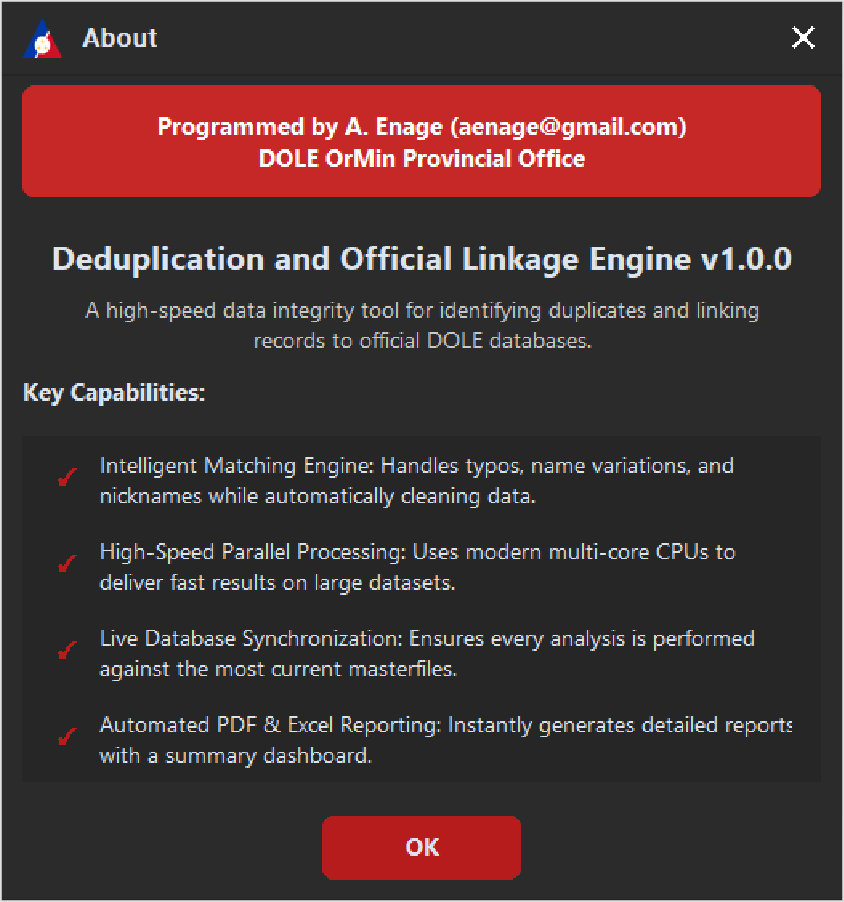
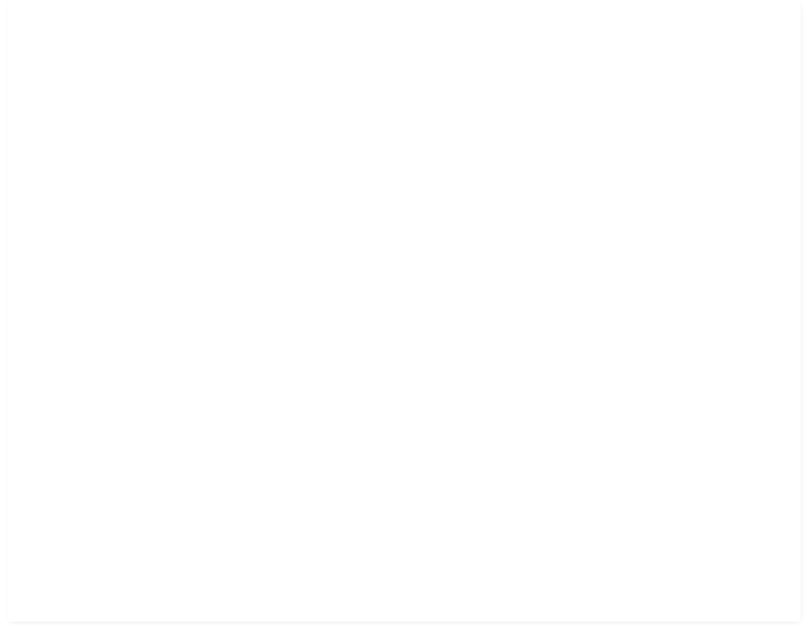
### Filename Format:

### Content: Flagged records requiring human verification

### Purpose: Quality assurance and final validation

### Usage: Manual review by authorized personnel for final decision-making

# Application Information and Credits



## About Dialog

*Figure 2.8*



Application Details: - **Developer:** A. Enage (aenage@gmail.com) - **Organization:** DOLE MIMAROPA Provincial Office - **Version:** {{VERSION}} – **Key Capabilities:** - Intelligent Matching Engine with nickname handling – High-Speed Parallel Processing for large datasets – Live Database Synchronization – Automated PDF & Excel Reporting

# Troubleshooting Guide

## Performance Optimization

First Run Considerations: - Initial startup may require up to 30 seconds for Windows component preparation – Subsequent launches will be significantly faster – Consider running from local drive rather than network location for optimal performance

## Common Issues and Solutions

### PDF Export Problems

**Issue:** PDF export fails or produces errors **Solutions:** - Verify Microsoft Excel (desktop version) is installed and properly licensed - Ensure Excel can open and save files without errors - Check for sufficient disk space in output directory - Restart Excel if it's currently running

### Data Import Issues

**Issue:** Column mapping errors or data not recognized **Solutions:** - Verify first row contains clear, descriptive column headers - Ensure all required columns are present and properly named - Check for special characters or unusual formatting in headers - Remove any merged cells or complex formatting from input file

### Network Connectivity Problems

**Issue:** Application cannot download reference files **Solutions:** - Verify internet connection is stable and active - Check firewall settings allow application internet access - Consider running initial setup on a different network if corporate restrictions apply - Use offline mode with cached files if internet is unavailable

### Cache and Performance Issues

**Issue:** Slow performance or outdated reference data **Solutions:** - Use "Clear Cache" option in Tools & Resources to refresh reference files - Ensure adequate disk space for cache storage - Restart application after clearing cache - Verify system date and time are correct for proper cache validation

## Data Quality Best Practices

**Input Data Preparation**: - Consistent Formatting: Use standardized date formats throughout your dataset - Complete Information: Ensure all required fields contain valid data - Clean Data: Remove extra spaces, special characters, and formatting - Standardized Names: Use consistent naming conventions for cities and locations

**Contact Number Handling:** - Numbers are preserved exactly as entered for follow-up purposes – No

automatic formatting or normalization is applied - Include country/area codes when available for

better tracking

# Support and Contact Information

## Technical Support

For technical assistance, application issues, or questions about functionality: - Contact: DOLE MIMAROPA administrator

## Reporting Issues

When contacting support, please provide: - Application version number - Operating system details - Description of the issue - Steps to reproduce the problem - Any error messages displayed

# Version Information

### Application Version: {{VERSION}}

* Release Date: {{DATE}}
* Developer: A. Enage (aenage@gmail.com)
* Copyright: © 2025 DOLE - All Rights Reserved

# Quick Reference Guide

## Essential Steps Summary Prepare Data - Format according to template requirements

1. Load File - Use Load button to import your dataset
2. Configure Settings - Access Tools & Resources for options
3. Run Analysis - Execute deduplication and matching process
4. Review Results - Examine generated reports and logs
5. Use Auditor - Perform additional quality assurance if needed

## Key Features

1. High-accuracy name matching with nickname equivalence
2. Phonetic and fuzzy matching algorithms
3. Government official detection in beneficiary lists
4. Comprehensive Excel and PDF reporting
5. Offline operation capability with caching
6. Dedicated auditor tool for quality assurance
7. No installation required - portable application