

Introduction

This section details how to use GMS Duplicate Person Handler, a tool for searching your database for people who exist more than once, then combining the duplicated records together.

Through the course of data entry there are many places where an person can be entered into GMS multiple times. Prior to GMS 5 you had the option to “dedup” those entrants you found that were duplicates one at a time. Now you have the option of running the Duplicate Person Handler to search and merge large numbers of entries.

After deduping (the process of merging duplicated records together and eliminating the excess copies), the resulting people will have all games, events, custom fields, certifications and addresses for both the source and target people.

Note on duplicated people

Once merged, the unique GMS ID field for the removed record is no longer valid, yet data in other databases may still have that person. Importing data from those systems using GMS Exchange will cause the duplicated person to be re-created, and so you may need to re-run this tool to eliminate them again. Alternately, you can run this tool on the data in other offices to eliminate the duplicates in all places.

Getting Started

From the main GMS screen, click on the “Tools” menu then click “Tools - Find and Handle Duplicate People” There are four phases to the process: Analyzer, Evaluate potential duplicates, Merge selected people, and Cleanup, and they *must* be done in this order to work correctly.

Phase 1: Analyzer

From the main Duplicate Person Manager window, click on “Phase 1”.

There are many options here for tuning and customization of the duplicate detection process.

Ignore possible duplicates...

If checked, GMS will not even consider that two people in different permanent delegations may be the same, even if their personal data is very close.

Minimum match score

Two people must accumulate at least this many points from the options below to be considered potential duplicates.

Duplicate Person Analyzer
This tool will analyze all of the people in your system to try to find potential duplicates - people with similar information that may be the same person, but are stored in GMS as separate people with different ID codes.

GMS will produce a list of potential duplicates. You will need to save this results file, then go back to the Duplication Manager and Evaluate each of the records and tell GMS whether or not each really is the same person.

Options

Ignore possible duplicates that are in different permanent delegations

Minimum match score:

Permanent delegations to work with:

Score Tuning Options

Name matches exactly: <input type="text" value="20"/>	DOB matches: <input type="text" value="15"/>
Last name matches exactly: <input type="text" value="10"/>	DOB matches 2 of 3 parts: <input type="text" value="6"/>
Last name is close: <input type="text" value="8"/>	DOB does not match: <input type="text" value="-8"/>
First name matches exactly: <input type="text" value="5"/>	Gender does not match: <input type="text" value="-15"/>
First name is close: <input type="text" value="3"/>	Soc sec# matches: <input type="text" value="20"/>

GMS 5, 5.0.1.15, Using back-end NexusDB, Using NexusDB server at address "127.0.0.1", NexusDB database "temp2", Logged in as wcornett.
Tables: 0/4 (created: 2), Locks: 0; Open lock tables: 0

Illustration 1, Analyzer options

Permanent delegations to work with

To analyze only a subset of your permanent delegations, check just those on the list you want to analyze. People outside of this set will not be considered for matches against those in this group.

Score Tuning Options

GMS will compare every person in the selected permanent delegations against every other person and assign each matchup a score based on what information they have in common. When one or more of the listed conditions is met, the matchup gets this value added to their score.

Notes on score tuning:

- Scores can be negative: the standard tuning options assign -15 to a matchup when the genders are different to indicate that even two people with the same date of birth are unlikely to be the same person if one is male and one is female.
- If one or more entrant has no date of birth, gender or Soc sec#, the associated rules are not evaluated. This means, for example, that when you compare a male to an “unknown”, GMS will not reduce the matchup's score.
- Setting the minimum score very low will result in more people to compare in Phase 2, but you can easily sort people by score there to eliminate the ones unlikely to be matches.

After setting up your options, click on [Execute]. This may take a long time to process if there are many people in your system, and it can be a good thing to run overnight, especially if your machine or network are slow.

When the analysis process is complete, GMS has come up with its list of potential duplicates. You need to save this file in order to continue. Type in the location and name of your search, or click the icon at the end of the line and save your search. Once you've named it, click the [Save Results] button (see Illustration 2, Save Analysis Results).

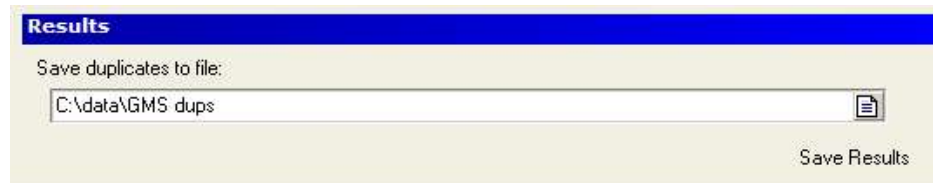


Illustration 2, Save Analysis Results

After saving your results, click on the [Back] to return to the Duplicate Person Manager main screen.

Phase 2: Evaluate potential duplicates

In this stage, you'll look at all of the record pairs that GMS examined and tell GMS whether or not they are really duplicates.

Click on “Phase 2: Evaluate potential duplicates”. Enter the name of the file you saved in the previous step (if you haven't closed GMS, its name will come up automatically), then click on [Open file].

Source name	Target name	Source delegation	Target delegation	Score	Status
Staff, Maryhaven #3	Staff, Maryhaven #1	Long Island Region	Long Island Region	20	Undecided
Staff, Maryhaven #3	Staff, Maryhaven #2	Long Island Region	Long Island Region	20	Undecided
Staff, AHRC_#2	Staff, AHRC #3	Long Island Region	Long Island Region	20	Undecided
Staff, Maryhaven #2	Staff, Maryhaven #1	Long Island Region	Long Island Region	20	Undecided
Pontecorvo, Brian	Pontecorvo, Brian	Long Island Region	Long Island Region	20	Undecided
Staff, AHRC #1	Staff, AHRC_#2	Long Island Region	Long Island Region	20	Undecided
Staff, AHRC #1	Staff, AHRC #3	Long Island Region	Long Island Region	20	Undecided
Green, Luis A.	Green, Shawn	Long Island Region	Long Island Region	16	Undecided
O'Rourke, Sean E	Ingbar, Marty	Long Island Region	Long Island Region	15	Undecided
Cranshaw, Jacob	Bellamy, Ashton	Long Island Region	Long Island Region	15	Undecided

View or edit this person <-> Swap source and target View or edit this person

Coach
Pontecorvo, Brian
DOB: n/a, Gender: Male
Long Island Region
Soc sec#:

(this person will have his/her data transferred to the first person and then be removed)

Athlete
Pontecorvo, Brian
DOB: 11/16/1982, Gender: Male
Long Island Region
Soc sec#:

(this person will be kept)

Status:

Illustration 3, Sample Duplicates

At the top you'll see a listing of all of the people which GMS considers potential duplicates. Clicking on any row brings up the details for these two people at the bottom of the screen where you can see more of their personal data, edit their personal data (by clicking on the “View or edit this person” links), choose which person of the two will be kept (by clicking on the [Swap source and target] button), and tell GMS whether or not these really are the same person.

Note: you can click on any of the columns in the listing to sort the people by that column. Normally the list is sorted in descending order of matchup score, which puts the most likely candidates to the top and the least likely candidates at the bottom.

Note: you can also right-click on any row and get a popup menu to let you change the matchup's state more quickly than using the drop-down at the bottom.

Source name	Target name	Source delegation	Target delegation	Score	Status
Pontecorvo, Brian	Pontecorvo, Brian	Long Island Region	Long Island Region	20	Undecided
Staff, AHRC	Staff, AHRC	Long Island Region	Long Island Region	20	Undecided
Staff, AHRC	Staff, AHRC	Long Island Region	Long Island Region	20	Undecided
Green, Luis	Green, Shawn	Long Island Region	Long Island Region	16	Undecided
O'Rourke, S	O'Rourke, S	Long Island Region	Long Island Region	15	Undecided
Cranshaw, Jacob	Bellamy, Ashton	Long Island Region	Long Island Region	15	Undecided

Yes - these are duplicates
No - these are not duplicates
Undecided

Illustration 4, Right-clicking on a matchup record

As you change the status of each matchup record, GMS saves the change to that record *but does not*

combine any people together. You can easily go back and make changes to these matchup records later, and you can make these comparisons over a period of multiple days.

Note: leaving a record marked as “Undecided” will cause GMS to not make any changes to these individuals later. If you find that matchup records with low matchup scores are not duplicates, you can just ignore them altogether rather than marking each and every one as “No – these are not duplicates”.

After checking all of the records you intend to, click on [Close file] then on the [Back] button to return to the Duplicate Person Manager.

Phase 3: Merge selected people

In this step, you'll tell GMS to commit the work you've done in the previous steps: GMS will take all of the records you've marked as duplicates and combine them together.

Click on “Phase 3: Merge selected people”. Enter the name of the file you saved in the previous step (if you haven't closed GMS, its name will come up automatically), then click on [Open file].

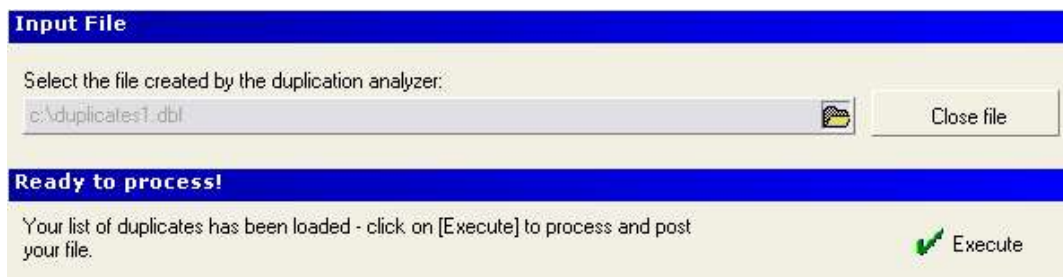


Illustration 5, loaded file ready to process

Note: once you run this phase, changes made to your data are *permanent*. Make a backup of your data prior to executing!

After the file loads, click the [Execute] button. You'll see the status bar at the bottom of the screen move toward 100%.

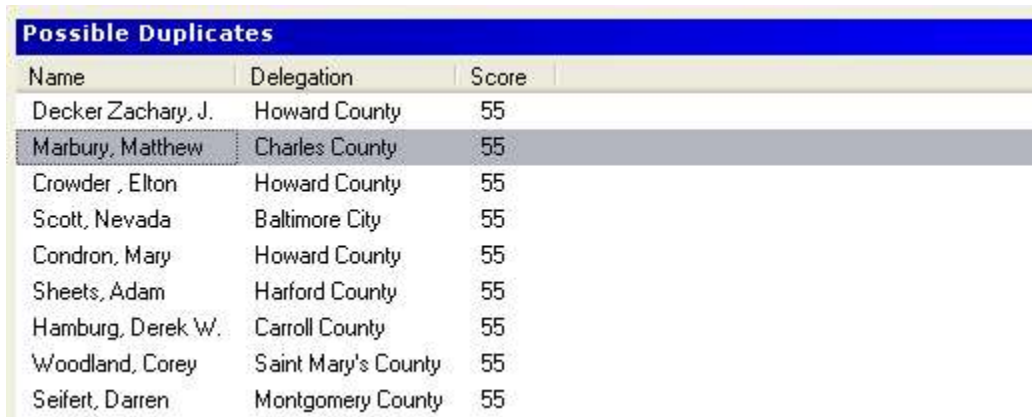
When finished, a print preview window will display all changes that were made to your data. You should probably print this report, either to a printer or a PDF file for later reference. After closing the preview window, you will get and you will see an option to save the processing results to a text file. (This file would have the same contents as the printed report previously run.) If you haven't printed the another section of the screen, giving you the option to save these results, open.

Click the [Back] to return to the Duplicate Person Manager.

Phase 4: Cleanup

This process gives you a list of the people who remain from the duplication process so you can edit them and verify their data. It is entirely optional – the duplicates are already gone.

Click on “Phase 4: Cleanup”. Enter the name of the file you saved in the previous step (if you haven't closed GMS, its name will come up automatically), then click on [Open file].



Possible Duplicates		
Name	Delegation	Score
Decker Zachary, J.	Howard County	55
Marbury, Matthew	Charles County	55
Crowder, Elton	Howard County	55
Scott, Nevada	Baltimore City	55
Condron, Mary	Howard County	55
Sheets, Adam	Harford County	55
Hamburg, Derek W.	Carroll County	55
Woodland, Corey	Saint Mary's County	55
Seifert, Darren	Montgomery County	55

Illustration 6, Cleanup list

To see or edit a person's data, double-click on his/her row and edit them as usual.