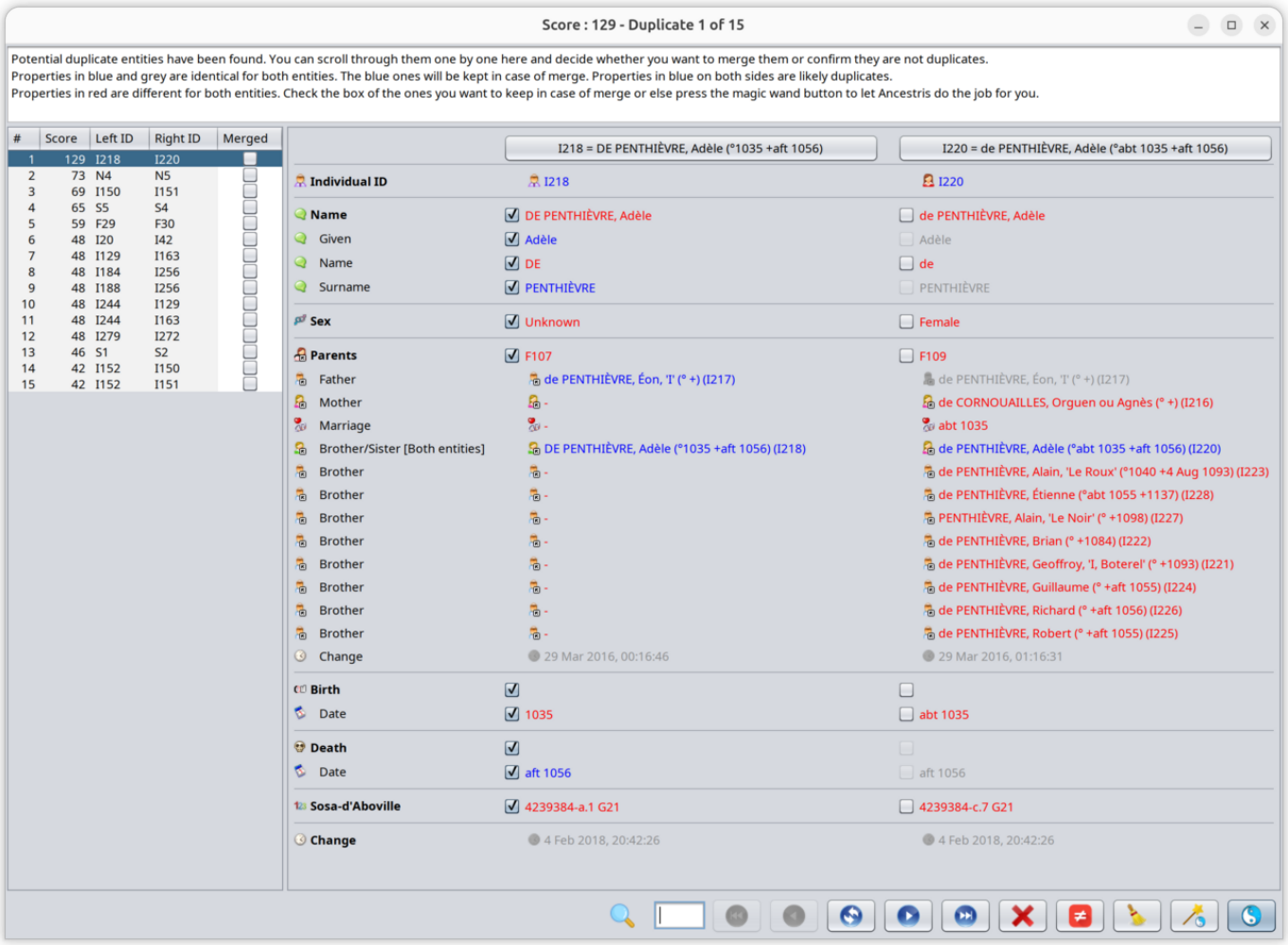
Merge duplicates

This tool detects duplicate entities and analyse them, merge them automatically or manually if they are real duplicates, or confirm them as non duplicates if they are not.

There are 4 ways to merge duplicates. They are described in this page.

- **Global search** across the whole genealogy file
- **Automatic detection** each time an entity is modified
- **Manual merge action** from the user to force the merge of two selected entities
- **Drag-and-Drop** or **Copy** of entities from one genealogy file to another

The following window is used in all situations.



Description

The Merge Duplicates tool is made of 3 components.

- **Entity selector window**: this is where you select which entities you want to search for duplicates in a global search.
- **Merge window**: this is where you will see all potential duplicate entities, analyse each of them and decide whether or not to merge them.
- **Special note for all non duplicates**: this is where Ancestris stores all your confirmations that two entities are not duplicates.

Before describing these 3 components, let's first introduce how Ancestris measures the likelihood of 2 entities being duplicates or not. Ancestris calculates a Ressemblance Score.

Resemblance Score

It is often difficult to assess with total certainty that two entities are duplicates or are not duplicates. Even a human person can sometimes have difficulty certifying that two individuals or entities are certainly the same or certainly not the same.

Of course, it would be easier to limit the detection to saying that two individuals with exactly the same surname, first name and date of birth are duplicates. In reality, dates could be either missing or be approximate, first names could be in different order or incomplete, etc. In these cases, you would still want Ancestris to be able to detect something.

> Therefore **Ancestris uses a resemblance score**. The higher the score, the more some information is similar, and the more likely the entities are duplicates.

The calculated scores can range from 0 to high positive numbers above 100. It is not a percentage.

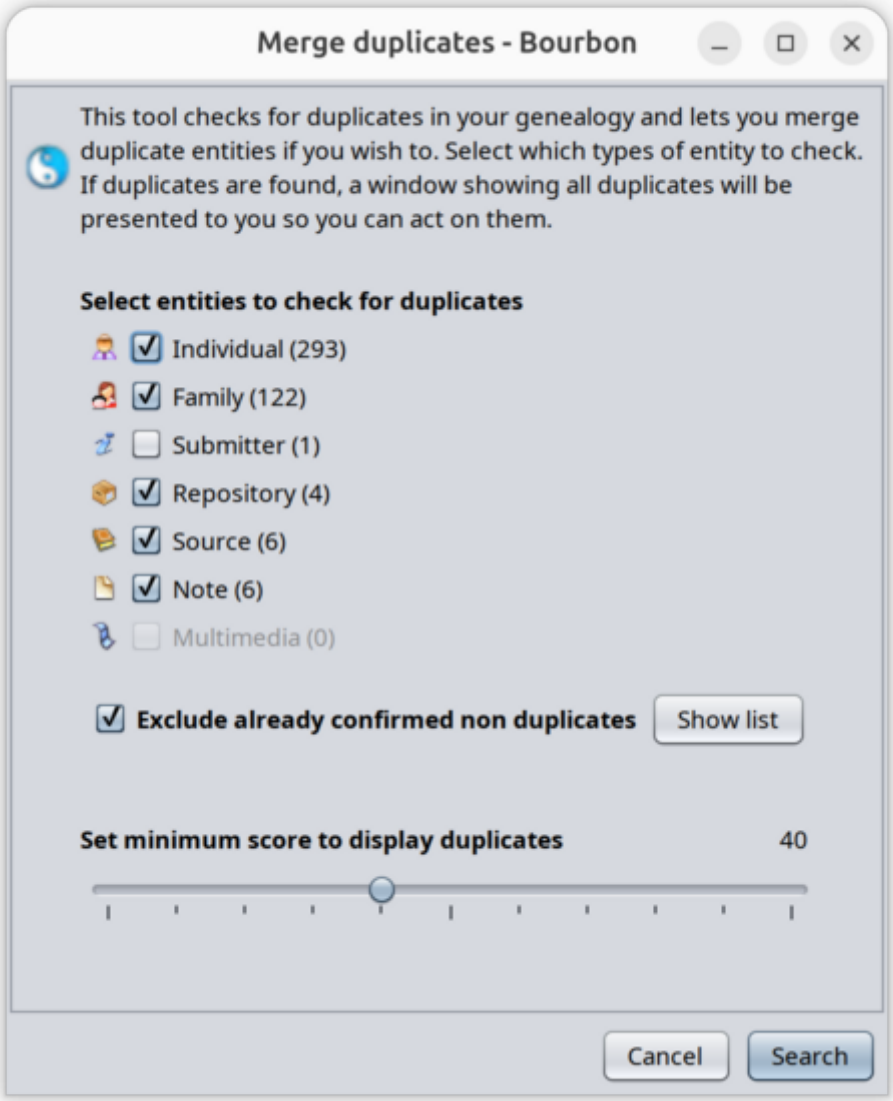
Ancestris will list the potential duplicates according to this indicator in a decreasing order, as an intent to tell you:"***While this is not certain, given the similarities in the information between these two individuals, they might be duplicates. And this is the score of confidence that they are***". Then it's up to you to merge the pair of entities or to discard the resemblance.

> What this means is that Ancestris might show you potential duplicates that you will consider as non duplicates (ex: twins with missing information), and reversely, Ancestris might miss to show you some duplicates that could actually be real duplicates according to you.
> Please accept our apologies if the detector is not perfect and please let us know if you find such instances.

> See below the **setting of the Resemblance score threshold**.

Entity selector window

When a global search for duplicates is launched, the following window is displayed. It shows the various types of entities and how many of each exist in the Gedcom file.



Entity selection

Check the entity boxes for which you want to search for duplicates.

Only the boxes of entities that are present in the Gedcom file are enabled.

In the example above, as there are no multimedia entities.

Known non-duplicates

In case you want to exclude known non duplicates from the search, i.e. those that you have already confirmed as non-duplicates, check the corresponding box.

To see the list of pairs of entities that you have confirmed as non duplicates, **press the "Show List" button.**

If you press the Show List button, the **list of Non-Duplicates** appear. It is described below.

Minimum score

A **scale of minimum score** can be used to help you exclude from the display scores which are to small.

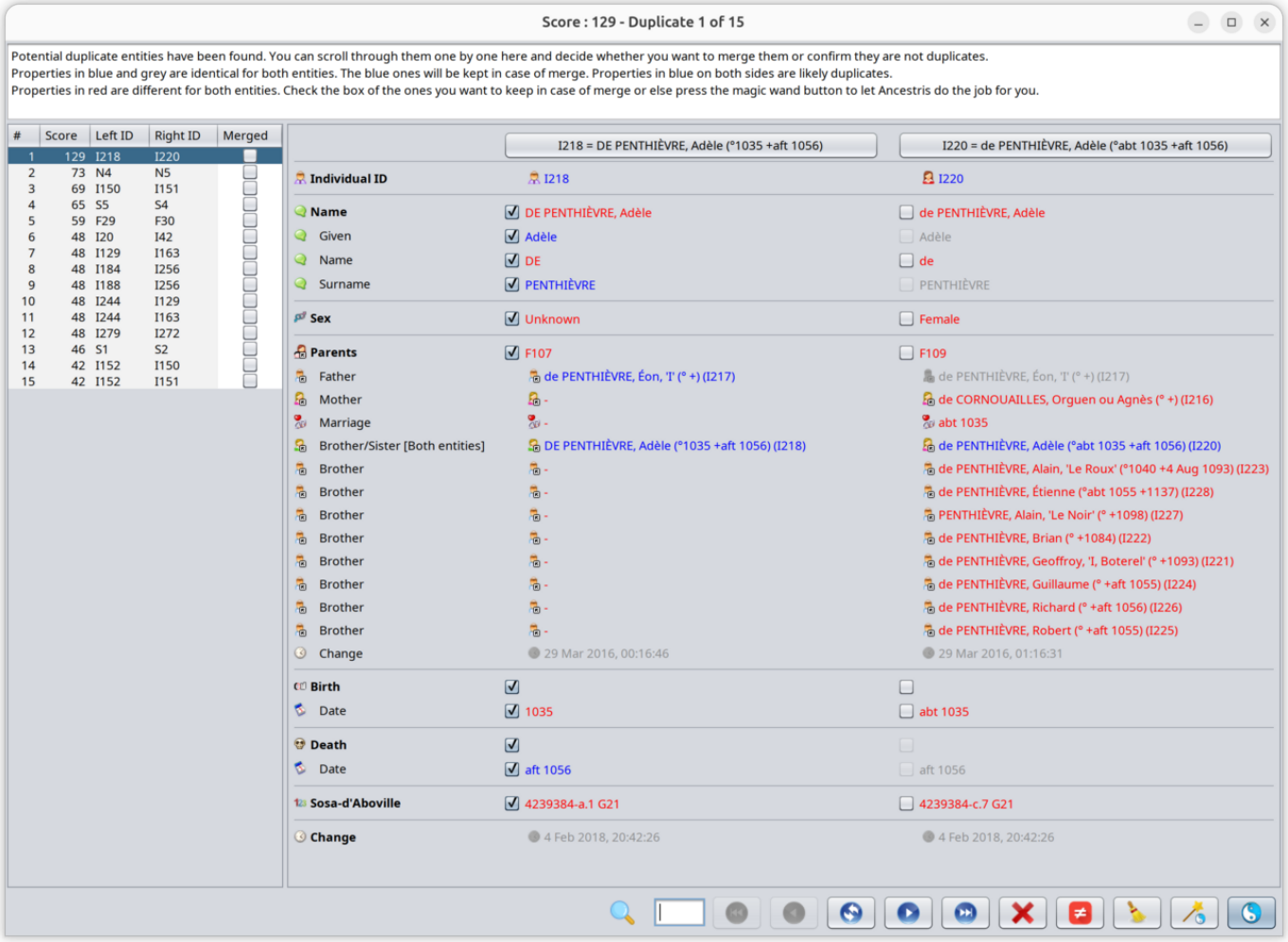
> Setting your own minimum score is useful when many potential duplicates are found, but the main reason it extists is to give you confindence that Ancestris found all potential duplicates and did not miss any.

A few words on how to set the minimum score.

- The ideal minimum threshold is the one that separates confirmed duplicates from confirmed non-duplicates.
- It does not exist in reality because some non real duplicate pairs can get a higher score than real ones in the vicinity of this threshold.
- This threshold depends on the genealogy of each user and is geneally between 40 or 50.
- You need to try several thresholds for your genealogy until you're satisfied that all duplicates displayed are mostly real duplicates, while still showing a few ones that are not, just to make sure you've got them all.

Merge window

For entities detected as duplicates, the following window is shown.



This window displays one by one the total **list of all potential duplicates where thescore is greater than the defined threshold**.
The list is sorted initialy **from the most certain pair of duplicates to the least certain** pair of duplicates, by category of entity. For each pair of similar entities, Ancestris displays the score indicator at the top.

Large genealogies can have several thousands of duplicates. This window will only display the first 10 000 duplicates for each entity type. If your genealogy has more duplicates, you will need to merge these before seing the others.

Title

The **title** of the window indicates the duplicate pair number displayed below, and the resemblance score that the two entities of this pair are in fact the same, and therefore to be merged.

Message

A **general message** is displayed at the top of the winodw. It depends on the situation in which the tool was launched, as global search, automatic detection, or manual action. This message also helps understand the color coding used to display the information.

Sortable selection list

On the left hand side is a sortable selection list of all pairs identified as potential duplicates.

You can sort the list by clicking on column headers.

Selecting the line of a potential duplicate will display the corresponding pair with the details in the right hand side of the window.

The '**merged**' flag indicates pairs of duplicates that you have merged using one of the buttons below.

Duplicate comparison details

Each pair of potential duplicates made of two entities is displayed in the two columns.

As title of each column, a button allows you to select each of the entities in the editors for more details.

In each column are displayed the properties of each entities of the supposed duplicate.

- Values that are different are displayed in red.
- Values that are identical are displayed in blue for the left hand side entity, in grey for the right hand side entity.
- Values that are likely duplicates are displayed in blue on both sides.

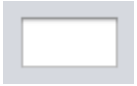
> The purpose of the comparison is to merge the right entity into the left one if you confirm them as duplicate

Therefore, a check box is available for each property on the right hand side to tell Ancestris to select manually what needs to be kept after merging them.

The buttons are used to navigate within the list of duplicate pairs, postpone the decision, merge them now or confirm now as non duplicate.

Button bar



Search duplicate field 

This field is used to search for duplicates in the list. Type the text to search in the entity names of the duplicate. Then press Enter. The Next and Previous buttons below can be use to search for the

next and previous match respectively.

Go to first duplicate Button 

Displays the first duplicate of the list in the current sort order.

Go to previous duplicate Button 

Displays the previous duplicate. If a search text exists in the Search duplicate field, it will display the previous duplicate matching this search criteria.

Swap Left and Right Entities Button 

Swap the left and right entities in order to merge the two entities on the left one and delete the right one. This is useful if most of the information to be kept after the merge is on the deleted entity.

Go to next duplicate Button 

Displays the next duplicate. If a search text exists in the Search duplicate field, it will display the next duplicate matching this search criteria.

Go to last duplicate Button 

Displays the last duplicate of the list in the current sort order.

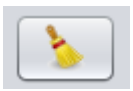
Close Button 

Closes the window.

Non duplicate Button 

This button **excludes the pair of entities from the potential duplicates**.

It marks the entity pair to be non duplicate and stores this confirmation in the **special "non duplicates" note**.

Remove duplicate Button 

Removes the potential duplicate from the displayed list.

It is useful if you do not know yet whether the two entities are duplicates or not, and you want to **postpone the decision**.

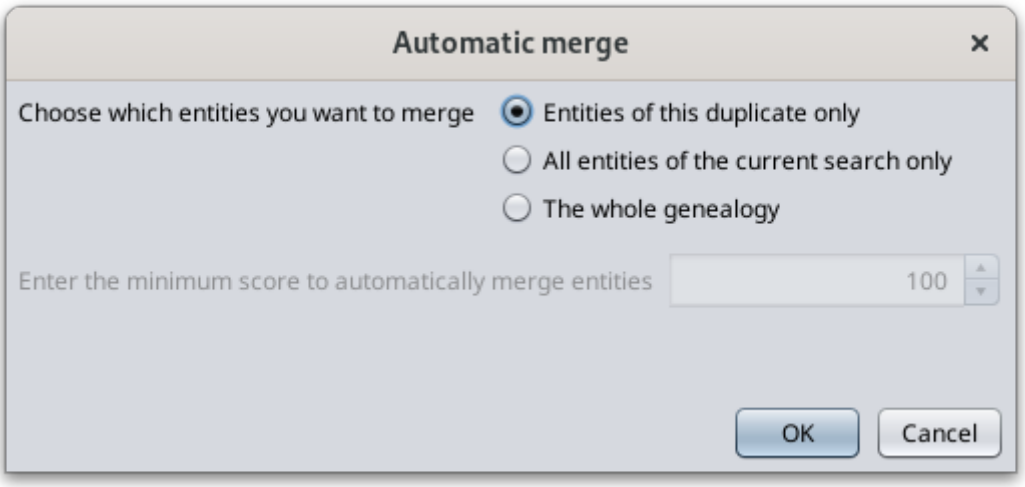
If a new global duplicate search is started, the duplicate will reappear.

Automatic Merge Button 

By clicking the Automatic Merge button, the **entities will be merged automatically and NOT using the check boxes.**

Ancestris will determine which information from the entity to be deleted should be kept and enrich the information of the entity to be kept.

Ancestris lets you perform this automatic merge for 3 different possible scopes. When you press the Automatic Merge Button, the following choices appear.



The choices are

- **Entities of this duplicate only:**
  - Only the current duplicate displayed will be merged.
  - Ancestris will detect which information to keep on the right hand side entity to enrich the left hand side entity.
  - Then the duplicate will update and show only the left entity in green text with the resulting information.
  - The Output window (Ctrl+T) will list the merge entities and the score. You can analyse this file and save it as reference.

- **All entities of the current search only:**
  - The automatic merge described above will be performed for all duplicates of the list with a score above a given score.You will need to specify that score in the field below.

- Only the first 10 000 duplicates of each entity will be considered. Use the choice described below if you want to merge the whole genealogy, not just the first 10 000 duplicates found.
- The merge window will remain open and will include all merged duplicates above the indicated score, as well as all the other duplicates.
- All duplicates including deleted entities will be removed from the list.
- The Output window (Ctrl+T) will list all merged entities. You can analyse and save this output file as reference.

- **The whole genealogy:**
  - The automatic merge described above will be performed for all duplicates of the genealogy above a given score, not just the first 10 000. You will need to specify that score in the field below.
  - The merge window will close.
  - The Output window (Ctrl+T) will list all merged entities. You can analyse and save this output file as reference.

Manual Merge Button 

By clicking the Manual Merge button, the **entities are merged using the check boxes**.

The entity on the right is removed from the Gedcom file and the information which check box is checked on the right hand side is added to the entity on the left or will replace the information on the left hand side.

For information that can only exist once, it is only possible to keep the information from one of the two entities.

As soon as the merge is done, the window displays the same duplicate in green with the result of the merge so that you can check that everything has been kept as you wanted.

You can then move on to the next duplicate.

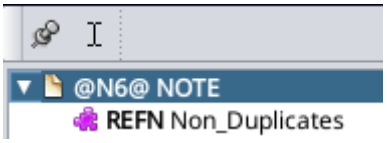
Special note for all non duplicates

> A special note is created and updated in Ancestris to store the non duplicate confirmations.

This note stores user confirmations of similar pairs that are actually not duplicates according to you.

It avoids Ancestris detecting them again and again each time the global search or the automatic detection is run.

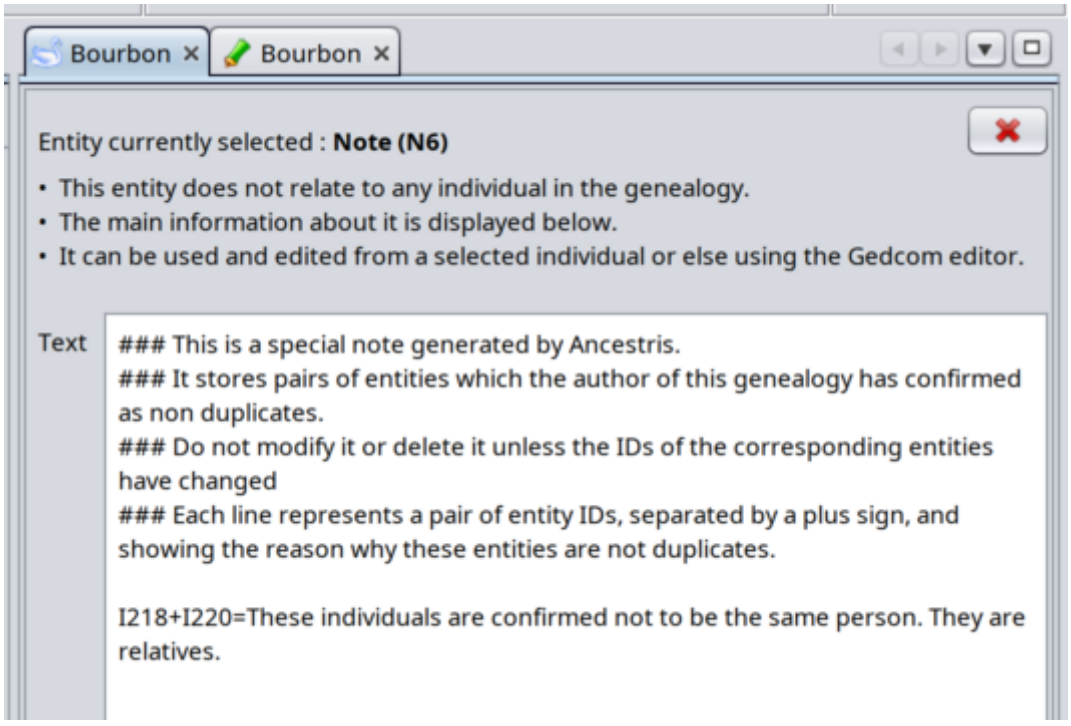
This note has a reference name called "Non_Duplicates".



The note is updated each time you press the Non Duplicate button in the **Merge window** or update the list using the **Entity selector window**.

> We have chosen to store this information in the Gedcom file itself because we value your efforts to analyse the entities and decide that they are not duplicates. We consider this information a valuable piece of genealogy information that has to be kept and transferred as part of the Gedcom file.
> The Gedcom standard does not cater for this need, hence Ancestris choice to store it this way, in one single note.
> Should the Gedcom standard evolve, such as a NOALIAS tag, we might change the way this information would be managed.

This note will appear as an isolated note. We recommend you do not delete it. The text in the note explains what this is.



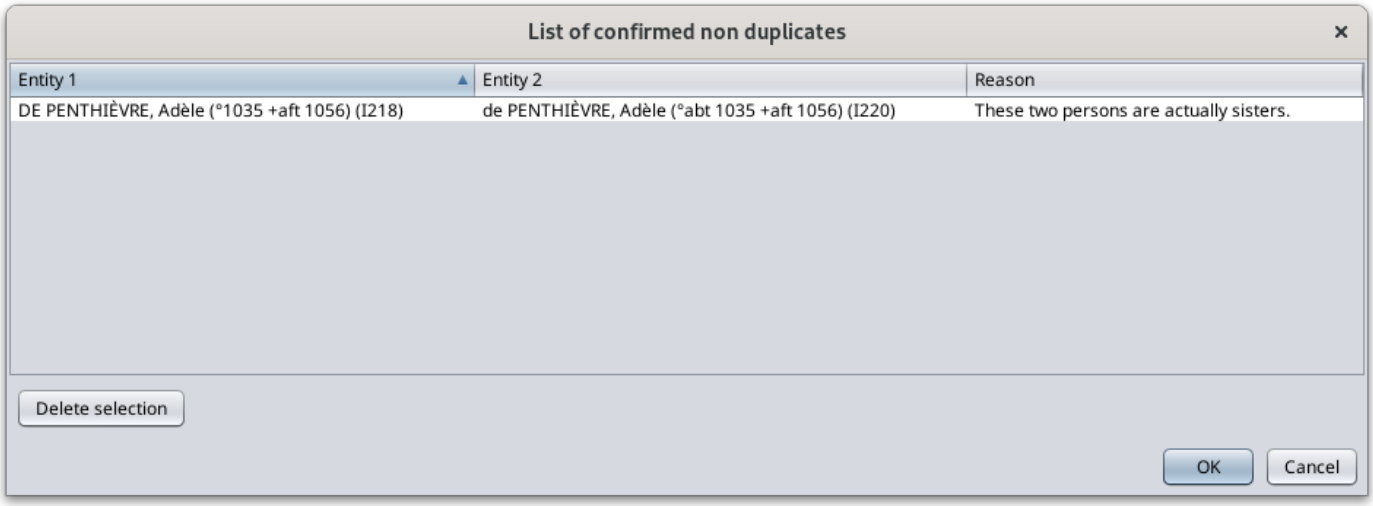
You can see the list of confirmed non duplicates from the **entity selector window**.

List of non-duplicates

You can see the list of non-duplicates by pressing the **"Show list" button on the Detection Criteria window**.

You can sort the lines to find the entities you are interested in.

You can select one or several lines and remove them from the list if you need to.



Usage

As mentioned above, there are 4 ways to merge entities.

- **Global search** across the whole genealogy file.
- **Automatic detection** each time an entity is modified.
- **Manual merge action** from the user to force the merge of two selected entities.
- **Drag-and-Drop** or **Copy** of entities from one genealogy file to another.

Global search

> The purpose of the global search is both to **identify duplicates throughout the whole genealogy and act on them**, that is decide one by one what you want to do with each them, or let Ancestris 'mass-merge' all duplicates above a given score.

Your decision for each duplicate will then be to either

1. **merge** the duplicate,
2. declare it as a **non duplicate**,
3. or **postpone** the decision to later.

You can launch the global search from the Ancestris tools menu.

The duplicate merge tool works in two steps.

- First you choose which entity type you want Ancestris to perform the detection of duplicates,
- Then you choose if and how to **merge duplicates** in the Merge window.

While using the tool, the genealogy is changed accordingly

1. Entities you decide to merge are merged with the information you specified to keep or let Ancestris choose to keep,
2. Entities you declare as non-duplicates are logged into the **special note**.

Automatic detection

> The purpose of the Automatic Detection is to alert you in case the entity you are currently creating or modifying is a potential duplicate of another entity already existing in your genealogy.

The automatic detection of duplicates is activated by default in the **Ancestris preferences**.

As soon as you validate your entry in one of the editors, and if the corresponding preference box is checked, the detection automatically searches potential duplicates of the entity being edited.

All potential duplicates are then presented in the **Merge window**, for you to decide what you want to do with these duplicates.

> In the case of the Cygnus editor where several entities can be edited at the same time (the individual , the family, the note, the source and repository related to an individual), all

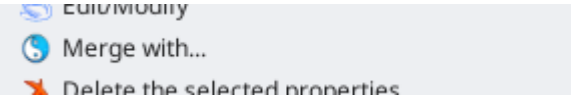
modified entities are checked and therefore Ancestris will list in the Merge window all the potential duplicates of all the modified entities.

Manual Merge action

The purpose of the Manual Merge action is to merge two entities, regardless of whether Ancestris detected them or not.

Another purpose is to identify any duplicate for a given entity.

This action is accessible from the **Context menu** on the current entity you want to merge with another one.



When this action menu item is selected, Ancestris asks you which other entity the current entity is to be merged with.

- **To really merge with another entity**, pick the one you think is a duplicate.
- **To just know whether the current entity has got duplicates in the genealogy**, just choose any entity from the list of entities.

Then Ancestris displays the Merge window with a list of potential duplicates among which will be the pair of two entities you chose and its corresponding score of being the same entity, which can be 0.

The list will be sorted in decreasing score and and Ancestris will position the selection on the pair of entities you chose at the start.

If the current entity you started from has no other found duplicates, only the pair of chosen entities will be shown in the list.

Then you may decide to merge the current entity you started from with the other chosen entity, or any other entity from the list.

Drag-and-Drop or Copy from a genealogy to another

> The purpose of the **Drag-and-Drop** / **Copy entities across genealogies** is to copy entities from one genealogy to another one using the mouse or the tools menu.

Customization

There are 2 customization information elements for the Merge tool.

- The **automatic detection flag**, stored in the **Preferences in the User Directory**.
- The **non-duplicate pairs**, stored in a special note.

Revision #51
Created 24 May 2020 17:54:43 by frederic
Updated 21 February 2026 17:19:51 by frederic