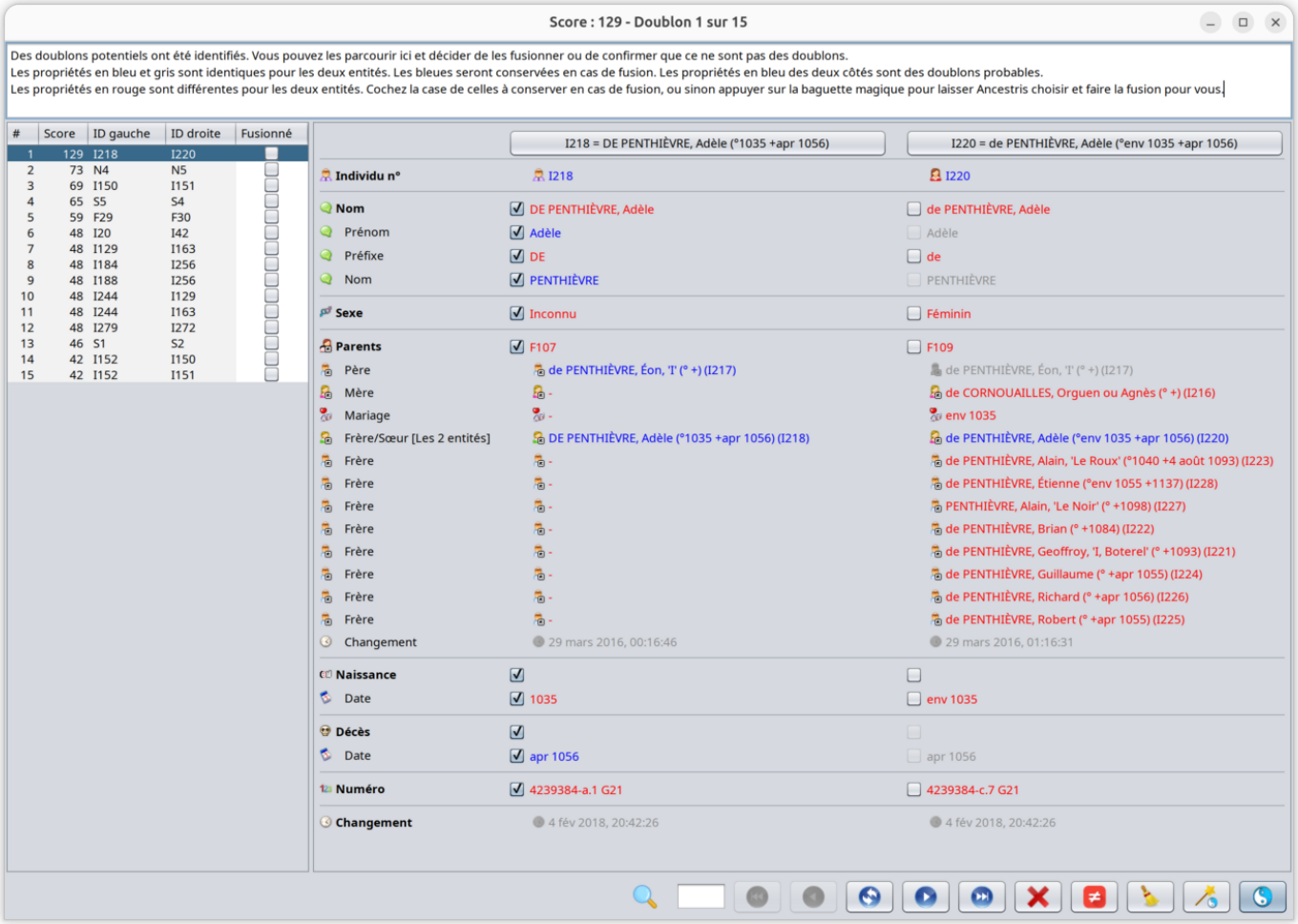
Fusion de doublons

Cet outil de Fusion des Doublons détecte les doublons et vous permet soit de les fusionner automatiquement ou manuellement s'il s'agit de vrais doublons, soit de les confirmer comme non doublons sinon.

Il existe quatre façons de rechercher et de fusionner les doublons. Elles sont expliquées dans cette page.

- **Recherche globale** à travers toute la généalogie
- **Détection automatique** à chaque fois qu'une entité est modifiée
- **Action manuelle de fusion** par l'utilisateur pour forcer la fusion de deux entités sélectionnées
- **Glisser-Déposer** ou **Copie** d'entités d'une généalogie à une autre

La fenêtre suivante est utilisée dans toutes ces situations.



Description

L'outil de Fusion des Doublons est composé de 3 éléments.

- **Fenêtre de sélection des entités**: c'est ici que vous choisissez le type d'entités pour la recherche des doublon dans toutes la généalogie.
- **Fenêtre de Fusion**: c'est ici que vous verrez tous les doublons potentiels et que vous déciderez de les fusionner ou non.
- **Note spéciale des non doublons**: c'est ici qu'Ancestris stocke toutes vos confirmations que deux entités ne sont pas des doublons.

Avant de décrire ces 3 éléments, présentons d'abord comment Ancestris mesure le score de ressemblance permettant de mesurer si 2 entités sont éventuellement des doublons ou pas.

Score de ressemblance

Il est souvent difficile d'affirmer avec une complète certitude que deux entités sont des doublons ou ne sont pas des doublons. Même un être humain peut parfois avoir des difficultés à certifier que deux individus ou entités sont certainement les mêmes ou certainement pas les mêmes.

Bien sûr, il serait plus facile de limiter la détection en disant que deux personnes ayant exactement le même nom, le même prénom et la même date de naissance sont des doublons. En réalité, les dates peuvent être manquantes ou approximatives, les prénoms peuvent être dans un ordre différent ou incomplets, etc. Dans ces cas, vous voudriez quand même qu'Ancestris puisse détecter quelque chose.

> C'est pourquoi **Ancestris utilise un Score de ressemblance**. Plus le score est élevé, plus cela veut dire que certaines informations sont similaires, donc plus il est probable que les entités soient des doublons.

Les scores calculés peuvent aller de 0 à des nombres positifs élevés supérieurs à 100. Il ne s'agit pas de pourcentage.

Ancestris dresse ensuite la liste des doublons potentiels en fonction de ce score dans l'intention de vous dire : "***Bien que cela ne soit pas certain, étant donné les similitudes dans les informations entre ces deux personnes, il est possible qu'elles soient des doublons***. ***Et voici le niveau de confiance qu'il y a entre eux***". Ensuite, c'est à vous de décider de les fusionner ou de rejeter la ressemblance.

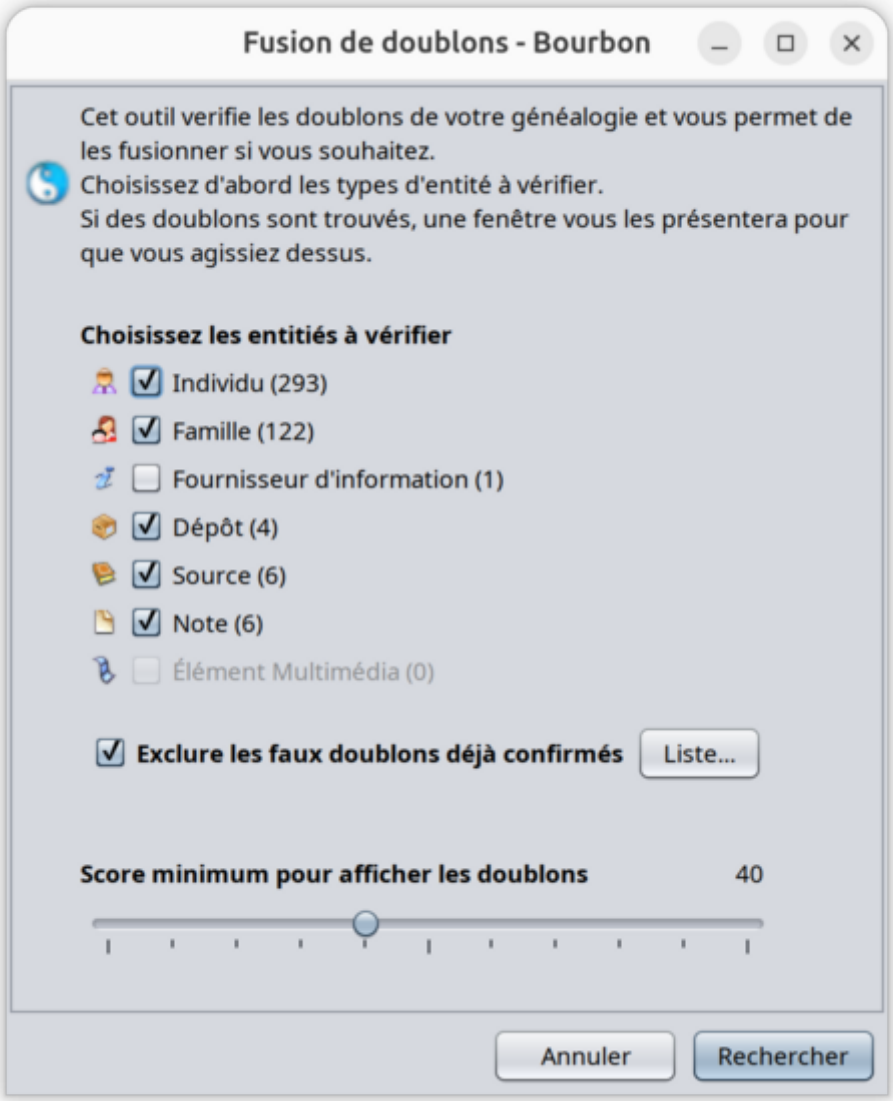
> Cela signifie qu'Ancestris peut vous montrer des doublons potentiels que vous considérerez comme non doublons, et inversement, Ancestris peut ne pas trouver certains doublons qui pourraient en fait être de vrais doublons selon vous.

> Veuillez accepter nos excuses si le détecteur n'est pas parfait et veuillez nous faire savoir si vous trouvez de tels cas.

> Voir ci-dessous le paramètre du seuil du score de ressemblance.

Fenêtre de sélection des entités

Lorsqu'une recherche globale de doublons est lancée, la fenêtre de sélection des entités s'affiche. Elle montre le nombre d'entités de chaque type et vous permet de les sélectionner pour lancer une recherche globale.



Sélection des entités

Cochez les cases des entités pour lesquelles vous souhaitez rechercher des doublons.

Seules les cases des entités présentes dans le fichier Gedcom sont activées. Dans l'exemple ci-dessus, comme il n'y a pas d'entités médias.

Non-doublons connus

Si **vous souhaitez exclure de la recherche les non-doublons connus**, c'est-à-dire ceux que vous avez déjà confirmés comme n'étant pas des doublons, cochez la case correspondante.

Pour voir la liste des paires d'entités que vous avez confirmées comme n'étant pas des doublons, **cliquez sur le bouton "Liste...".**

Si vous appuyez sur le bouton "Liste...", la **liste des entités non doublons** s'affiche.

Score minimum

Un **curseur de score minimum** à afficher peut être utilisée pour vous aider à exclure les scores trop faibles.

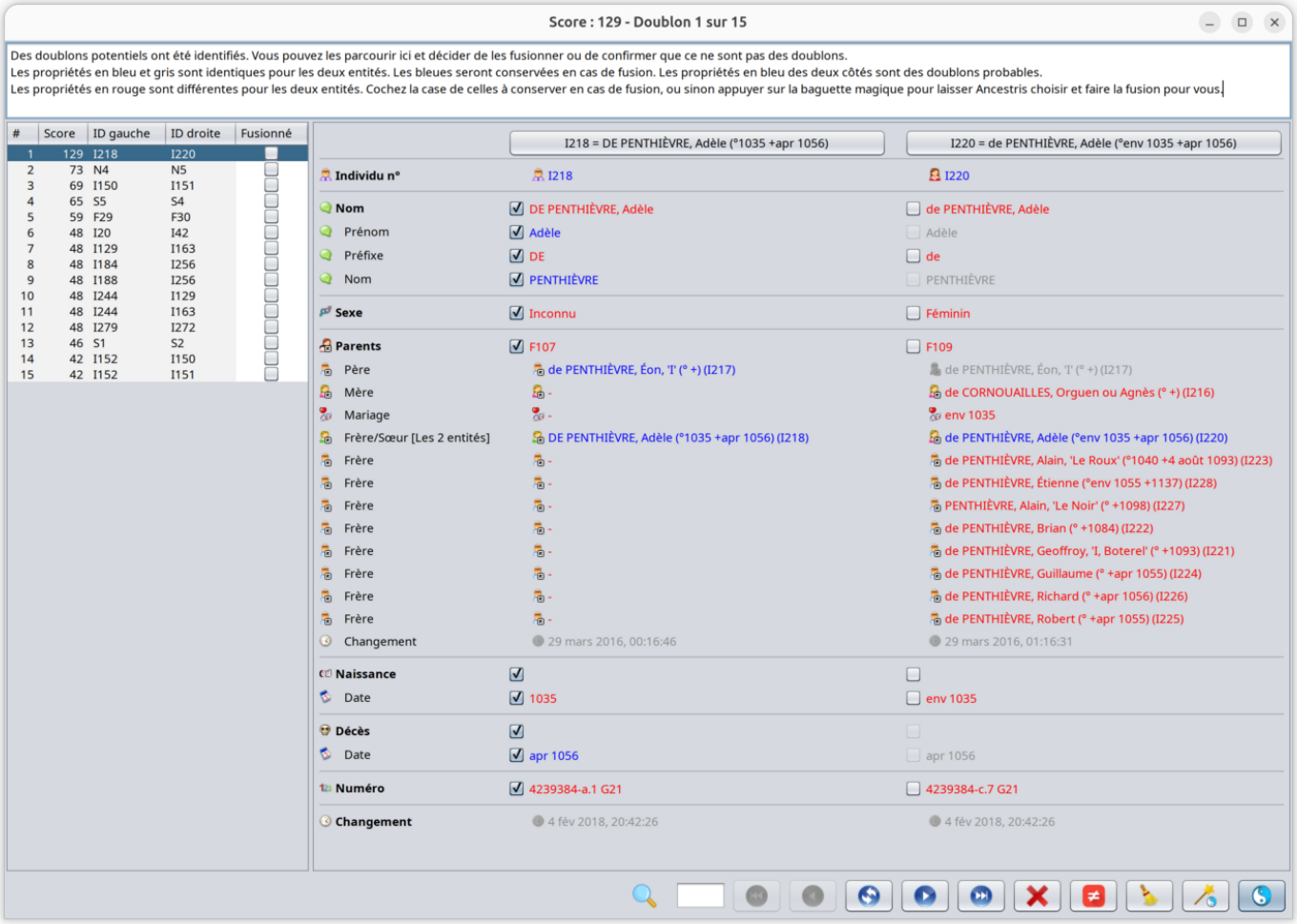
> Définir votre propre score minimum est utile lorsque de nombreux doublons potentiels sont trouvés, mais la raison principale de son existence est de vous donner l'assurance qu'Ancestris a bien trouvé tous les doublons potentiels et n'en a manqué aucun.

Quelques mots sur la façon de définir le score minimum.

- Le seuil minimum idéal est celui qui sépare les doublons confirmés des non-doublons confirmés.
- Il n'existe pas dans la réalité, car certaines paires de doublons non réels peuvent obtenir un score plus élevé que les doublons réels à proximité de ce seuil.
- Ce seuil dépend de la généalogie de chaque utilisateur et se situe généralement entre 40 et 50.
- Vous devez essayer plusieurs seuils pour votre généalogie jusqu'à ce que vous soyez satisfait que tous les doublons affichés soient pour la plupart de vrais doublons, tout en affichant encore quelques-uns qui ne le sont pas, juste pour vous assurer que vous les avez tous.

Fenêtre de Fusion

Pour les entités comparées en tant que doublons, la fenêtre suivante est utilisée.



> Cette fenêtre affiche un par un la **liste totale de tous les doublons potentiels dont la probabilité est supérieure au seuil choisi.**
> La liste est triée initialement **de la paire de doublons la plus certaine à la paire de doublons la moins certaine**, par catégorie d'entité. Pour chaque paire d'entités similaires, Ancestris vous donne un score de probabilité.

> Les grandes généalogies peuvent comporter plusieurs milliers de doublons. Cette fenêtre n'affichera que les 10 000 premiers doublons pour chaque type d'entité. Si votre généalogie comporte plus de doublons, vous devrez les fusionner avant de pouvoir voir les autres.

Titre

Le **titre de** la fenêtre indique le numéro de la paire de doublons affichée et le score de ressemblance que les deux entités de cette paire soit bien les mêmes, et donc à fusionner.

Message

Un **message général** est affiché et dépend de la situation dans laquelle l'outil a été lancé : recherche globale, détection automatique ou action manuelle.

Liste de sélection triable

Sur le côté gauche se trouve une liste de sélection triable de toutes les paires identifiées comme doublons potentiels.

Vous pouvez trier la liste en cliquant sur les en-têtes de colonne.

En sélectionnant la ligne d'un doublon potentiel, la paire correspondante s'affiche avec les détails dans la partie droite de la fenêtre.

La case « Fusionné » indique les paires de doublons que vous avez fusionnées à l'aide de l'un des boutons fusion situés en dessous.

Détails comparatif du doublon

Chaque paire de doublons constituée de deux entités est affichée dans les deux colonnes.

En titre de chaque colonne, un bouton permet de sélectionner chacune des entités dans les éditeurs pour plus de détails.

Dans chaque colonne sont affichées les propriétés de chacune des entités du doublon supposé.

- Les valeurs qui sont différentes sont affichées en rouge.
- Les valeurs identiques sont affichées en bleu pour l'entité de gauche, en gris pour l'entité de droite.
- Les valeurs qui se ressemblent sont affichées en bleu dans les deux colonnes.

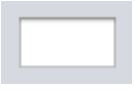
> L'objectif de la comparaison est de fusionner l'entité de droite dans celle de gauche si vous confirmez qu'elles sont dupliquées.

Par conséquent, une case à cocher est disponible pour chaque propriété sur le côté droit pour indiquer à Ancestris de conserver les informations sélectionnées de chaque entité après les avoir fusionnées.

Les boutons permettent de naviguer dans la liste des doublons potentiels, de reporter la décision, de les fusionner maintenant ou de confirmer maintenant qu'elles ne sont pas en double.

Barre de boutons



Champ Recherche de doublon 

Ce champ permet de rechercher un doublon dans la liste. Saisissez le texte et appuyez sur Entrée. Utilisez ensuite les boutons Précédent et Suivant pour se positionner sur le suivant ou le prochain doublon correspondant.

Bouton aller au début 

Affiche le premier doublon de la liste selon le tri en cours.

Bouton aller au précédent 

Affiche le doublon précédent. Si un texte de recherche existe dans le champ Rechercher de doublon, il affichera le doublon précédent correspondant à ce critère de recherche.

Bouton retourner les entités 

Intervertit les entités de gauche et de droite afin de fusionner ensuite les deux entités sur celle de droite. Cette opération est utile si la plupart des informations à conserver après la fusion se trouvent sur le côté droit.

Bouton aller au suivant 

Affiche le doublon suivant. Si un texte de recherche existe dans le champ Rechercher de doublon, il affichera le doublon suivant correspondant à ce critère de recherche.

Bouton aller au dernier 

Affiche le dernier doublon de la liste dans l'ordre de tri en cours.

Bouton Fermer 

Ferme la fenêtre.

Bouton Non doublon 

Ce bouton **exclut la paire d'entités des doublons potentiels**.

Il marque la paire d'entités comme n'étant pas en double et enregistre cette confirmation dans la **note spéciale "non duplicates"**.

Bouton retirer ce doublon de la liste 

Supprime le doublon potentiel de la liste affichée.

Cette option est utile si vous ne savez pas encore si les deux entités sont des doublons ou non, et si vous souhaitez **reporter la décision**.

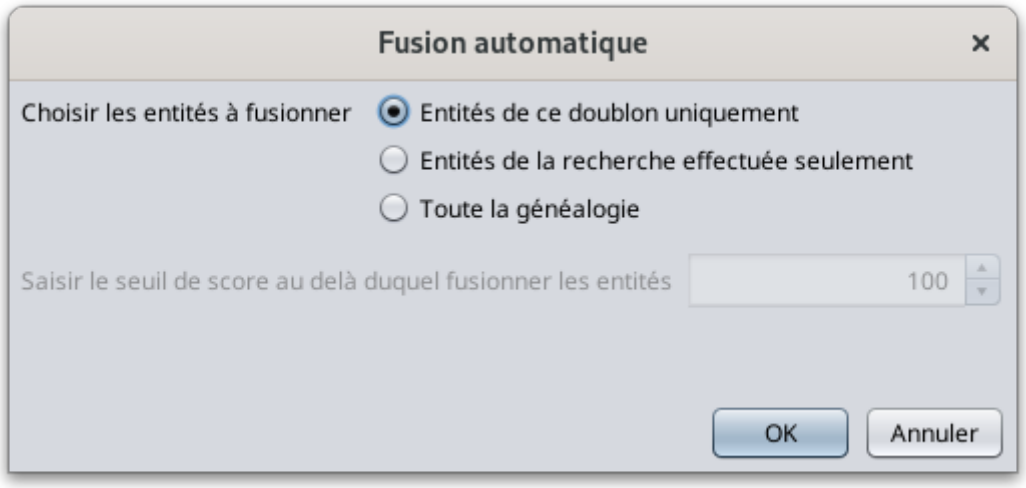
Si une nouvelle recherche globale de doublons est lancée, le doublon réapparaîtra.

Bouton Fusionner automatiquement 

En cliquant sur le bouton Fusionner, les **entités seront fusionnées automatiquement par Ancestris sans utiliser les cases à cocher**.

Ancestris déterminera quelles informations de l'entité à supprimer doivent être conservées et enrichira les informations de l'entité à conserver.

Ancestris vous permet d'effectuer cette fusion automatique pour 3 périmètres différents. Lorsque vous appuyez sur le bouton Fusion automatique, les choix suivants apparaissent.



Les choix sont les suivants

- **Entités de ce doublon uniquement :**
  - Seul le doublon actuellement affiché sera fusionné.
  - Ancestris détectera les informations à conserver sur l'entité de droite pour enrichir l'entité de gauche.
  - Ensuite, le doublon sera mis à jour et n'affichera que l'entité de gauche en texte vert avec les informations résultantes.
  - La fenêtre de sortie (Ctrl+T) affiche les entités fusionnées et le score. Vous pouvez analyser ce fichier et le sauvegarder comme référence.

- **Entités de la recherche effectuée uniquement :**
  - La fusion automatique décrite ci-dessus sera effectuée pour tous les doublons de la liste dont le score est supérieur à un certain seuil, que vous devrez spécifier dans le champ en dessous.
  - Seuls les 10 000 premiers doublons de chaque entité seront pris en compte. Utilisez le choix décrit ci-dessous si vous souhaitez fusionner toute la généalogie, et pas seulement les 10 000 premiers doublons trouvés.
  - La fenêtre de fusion restera ouverte et affichra tous les doublons fusionnés au-dessus du score indiqué, ainsi que tous les autres doublons non fusionnés.
  - Tous les doublons comprenant des entités supprimées seront supprimés de la liste.
  - La fenêtre de sortie (Ctrl+T) indiquera toutes les entités fusionnées. Vous pouvez analyser et enregistrer ce fichier de sortie comme référence.

- **Toute la généalogie :**
  - La fusion automatique décrite ci-dessus sera effectuée pour tous les doublons de la généalogie au-delà d'un score donné, et pas seulement pour les 10 000 premiers. Vous devrez spécifier ce score dans le champ en dessous.
  - La fenêtre de fusion se fermera.
  - La fenêtre de sortie (Ctrl+T) affichera la liste de tous les doublons trouvés.

Bouton Fusionner à la main 

En cliquant sur le bouton Fusionner, les **entités seront fusionnées en utilisant les cases cochées**.

L'entité de droite sera supprimée du fichier Gedcom et les informations dont la case est cochée à droite seront ajoutées à, ou bien remplaceront, celles de l'entité de gauche.

Pour les informations qui ne peuvent exister qu'une seule fois, il n'est possible de conserver les informations que dans l'une des deux entités.

Dès que la fusion est effectuée, la fenêtre affiche le même doublon avec le résultat de la fusion afin que vous puissiez vérifier que tout a été conservé comme vous le souhaitiez.

Vous pouvez ensuite passer au doublon suivant.

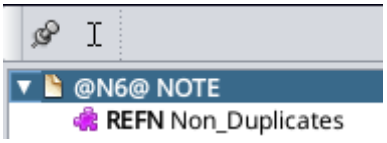
Note spéciale pour les non doublons

> Une note spéciale est créée et mise à jour dans Ancestris pour conserver les justifications des non-doublons confirmés.

Cette note conserve vos confirmations de paires similaires qui ne sont en fait pas des doublons selon vous.

Cela évite à Ancestris de les détecter encore et encore à chaque fois que la recherche globale ou la détection automatique est lancée.

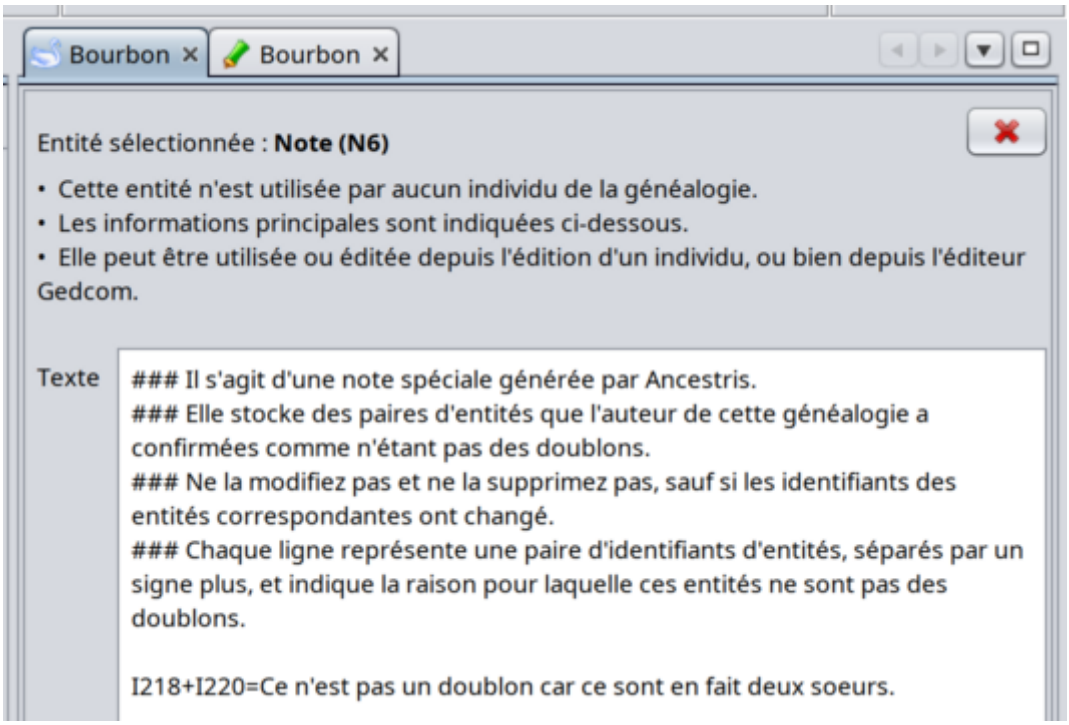
Cette note a un nom de référence appelé "Non_Duplicates".



Cette note est mise à jour chaque fois que vous appuyez sur le bouton "Non Doublon" dans la **fenêtre Fusion** ou que vous mettez à jour la liste à l'aide de la **fenêtre de sélection des Entités**.

> Nous avons choisi de conserver ces informations dans le fichier Gedcom lui-même parce que nous valorisons vos efforts d'analyse et décisions. Votre justification que deux entités ne sont pas des doublons constituent une information généalogique précieuse qui doit être conservée et transférée en tant que faisant partie du fichier Gedcom.
> La norme Gedcom ne répondant pas à ce besoin, le choix d'Ancestris est de la conserver de cette manière, dans une note unique.
> Si la norme Gedcom venait à évoluer sur ce point, par exemple avec une balise NOALIAS, nous pourrions alors changer la façon dont ces informations sont gérées.

Cette note apparaîtra comme une note isolée. Nous vous recommandons de ne pas la supprimer. Le texte de la note explique de quoi il s'agit .



Vous pouvez consulter la liste des non doublons confirmés dans la **fenêtre de sélection des Entités**.

Liste des Non-Doublons

Vous pouvez voir la liste des entités non doublon en cliquant sur le bouton **"Afficher la liste" dans la fenêtre Critères de détection**.

Vous pouvez trier les lignes pour trouver les entités qui vous intéressent.

Vous pouvez sélectionner une ou plusieurs lignes et les supprimer de la liste si nécessaire.



Usage

Comme indiqué plus haut, il existe trois façons d'utiliser cet outil.

- **Recherche globale** sur l'ensemble du fichier généalogique
- **Détection automatique** à chaque modification d'une entité
- **Action de fusion manuelle** par l'utilisateur pour forcer la fusion de deux entités sélectionnées
- **Glisser-Déposer** ou **Copie** d'entités d'une généalogie à une autre

Recherche globale

>

> Le but de la recherche globale est à la fois d'**identifier les doublons dans l'ensemble de la généalogie et d'agir sur eux**, c'est-à-dire de décider un par un ce que vous voulez faire sur chacun d'eux, ou laisser Ancestris effectuer une fusion d'ensemble pour vous au delà d'un score donné de ressemblance.

Pour chaque doublon, vous pourrez alors décider de

1. **fusionner** le doublon
2. déclarer qu'**il ne s'agit pas d'un doublon**,
3. ou de **reporter** la décision à plus tard.

Vous pouvez lancer la recherche globale à partir du menu Outils d'Ancestris.

L'outil de fusion des doublons fonctionne en deux étapes.

- Tout d'abord, vous spécifiez les **entités** dans la fenêtre correspondante,
- Ensuite, vous choisissez comment **fusionner les doublons** dans la fenêtre Fusionner.

Lors de l'utilisation de l'outil, la généalogie est modifiée en conséquence

1. Les entités que vous décidez de fusionner sont fusionnées avec les informations que vous avez spécifiées de conserver.
2. Les entités que vous déclarez non dupliquées sont enregistrées dans la **note spéciale**.

Détection automatique

> Le but de la détection automatique est de vous alerter au cas où l'entité que vous êtes en train de créer ou de modifier est un doublon potentiel d'une autre entité déjà existante dans votre généalogie.

La détection automatique des doublons est activée par défaut dans les **préférences d'Ancestris**.

Dès que vous validez votre saisie dans l'un des éditeurs, et si la case de préférence correspondante est cochée, la détection recherche automatiquement les doublons potentiels de l'entité en cours d'édition.

Tous les doublons potentiels sont alors présentés dans la **fenêtre Fusionner**, pour que vous puissiez décider de ce que vous voulez faire de ces doublons.

>

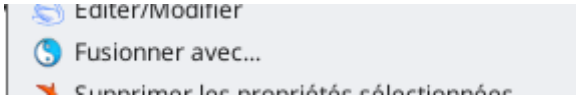
> Dans le cas de l'éditeur Cygnus où plusieurs entités peuvent être éditées en même temps, toutes les entités modifiées sont vérifiées et Ancestris listera donc dans la fenêtre Fusion tous les doublons potentiels de toutes les entités modifiées.

Action de fusion manuelle

> L'objectif de l'action Fusion manuelle est de fusionner deux entités, qu'Ancestris les ait détectées ou non.

> Un autre objectif est d'identifier tous les doublons d'une entité donnée.

Cette action est accessible depuis le **menu contextuel** de l'entité en cours et que vous souhaitez fusionner avec une autre.



Lorsque cette action du menu contextuel est sélectionnée, Ancestris vous demande avec quelle autre entité l'entité actuelle doit être fusionnée.

- **Pour fusionner réellement avec une autre entité**, choisissez celle que vous pensez être un doublon.
- **Pour seulement savoir si l'entité courante a des doublons dans la généalogie**, il suffit de choisir n'importe quelle entité dans la liste des entités.

Ancestris affiche alors la fenêtre Fusionner avec une liste de doublons potentiels parmi lesquels se trouve la paire de deux entités que vous avez choisie et sa probabilité correspondante d'être la même entité, qui peut être de 0%.

La liste sera triée par probabilité décroissante et la paire de doublons affichée sera la paire d'entités que vous avez choisie.

Si l'entité dont vous êtes parti n'a pas d'autres doublons trouvés, seule la paire d'entités choisie sera affichée dans la liste.

Vous pouvez alors décider de fusionner l'entité actuelle dont vous êtes parti avec l'autre entité choisie, ou avec toute autre entité de la liste.

Glisser-Déposer ou Copie d'une généalogie à une autre

> Le but du **glisser-déposer** / **copie d'entités entre généalogies** est de copier des entités d'une généalogie à une autre à l'aide de la souris ou du menu outils

Personnalisation

Il existe 2 paramètres de personnalisation pour l'outil de fusion.

- L'**indicateur de détection automatique**, stocké dans les **préférences du répertoire utilisateurs**.
- Les **paires non doublon**, stockées dans la note spéciale.

Revision #22
Created 29 March 2020 19:23:18 by frederic
Updated 21 February 2026 17:19:48 by frederic