

AFO 114 – Dédoublonnage de notices

114.1 Introduction

La fonction « dédoublonnage » permet de détecter des notices bibliographiques et autorités éventuellement identiques dans le catalogue. La fonction « fusion de notices » permet de consolider les éventuels doublons en une seule notice. Le but de ces opérations est donc de conserver l'homogénéité et la « propreté bibliographique » du catalogue.

La possibilité de fusion de notices dépend en premier lieu d'un diagnostic de doublons. Ce diagnostic peut être fait soit à la suite d'une simple constatation de catalogueur, à la suite de recherches dans la base de données (AFO 111 et AFO 113), soit à la suite de la mise en route d'un processus de recherche de doublons.

Dans le premier cas on s'orientera alors vers un processus de fusion manuelle, fonction disponible en AFO 111 et 113.

Dans le deuxième cas il est nécessaire de passer par une étape de création d'un fichier de doublons potentiels à partir de critères paramétrables.

Les AFOs concernées

- Le menu « Catalogage » (AFO 111 et 113)

- Le menu « Fusion des notices » (AFO 115)
 - Fonction « Fichiers de fusion »

 - Fonction « Profils de fusion »

- Le menu « Dédoublonnage » (AFO 114)
 - Fonction « Fichiers de dédoublonnage »

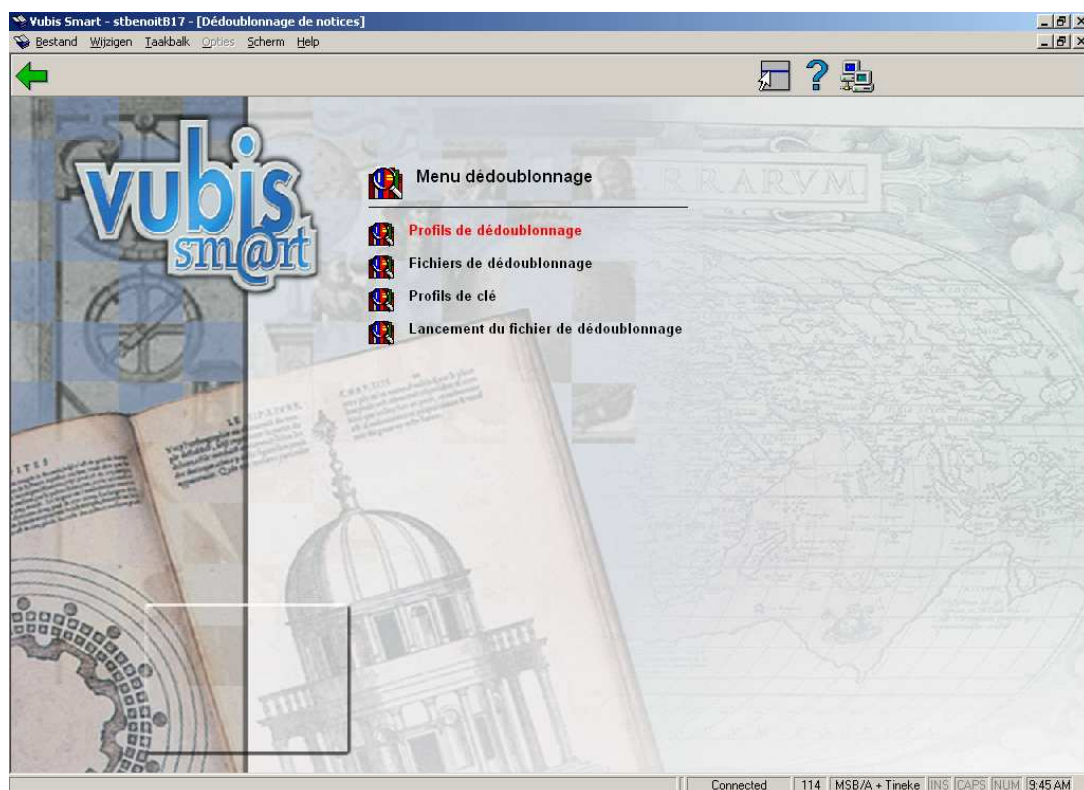
 - Fonction « Profils de clés »

 - Fonction « Lancement du fichier de dédoublonnage ».

Pour effectuer une recherche de doublons :

- 1. Premièrement, il faut créer un profil de clé
- 2. Deuxièmement, il faut créer un fichier de dédoublement
- 3. Pour finir, il faut lancer le fichier de dédoublement.

Lorsqu'on accède à l'AFO on obtient le menu suivant composé de plusieurs options :

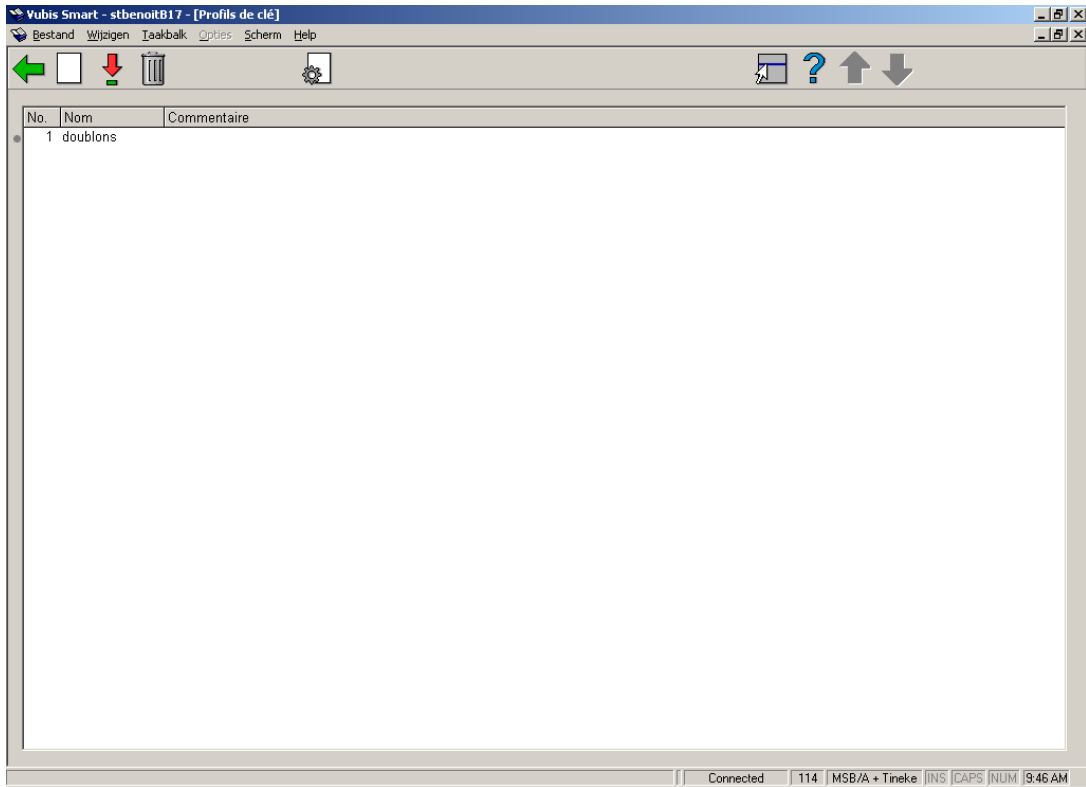


Les options du menu seront expliquées séparément dans les paragraphes suivants.

114.2 Profils de clé

Une clé comprend un ensemble d'éléments d'une notice. Il s'agit d'une chaîne de caractères qui ressemble à "financial^jone^elsev^2004" (extrait d'un exemple). Les éléments de la clé sont dérivés de la notice, dans l'exemple : le premier mot du titre, les quatre premiers caractères de l'auteur, les cinq premiers caractères de l'éditeur et l'année de publication. Puisque les éléments d'une clé peuvent provenir de zones répétables (comme les auteurs), plusieurs clés par titre sont possibles. Par comparaison, toutes les clés disponibles sont utilisées.

Après la sélection de cette option, un écran sommaire s'affichera :



Options sur l'écran

Nouveau: Après la sélection de cette option, un écran de saisie s'affichera :



Zones sur l'écran

Nom : nom du profil de clé.

Format : cliquez sur la flèche vers le bas pour sélectionner le format de l'enregistrement pour votre nouveau profil de clé.

Commentaire : texte qui décrit le profil de clé.

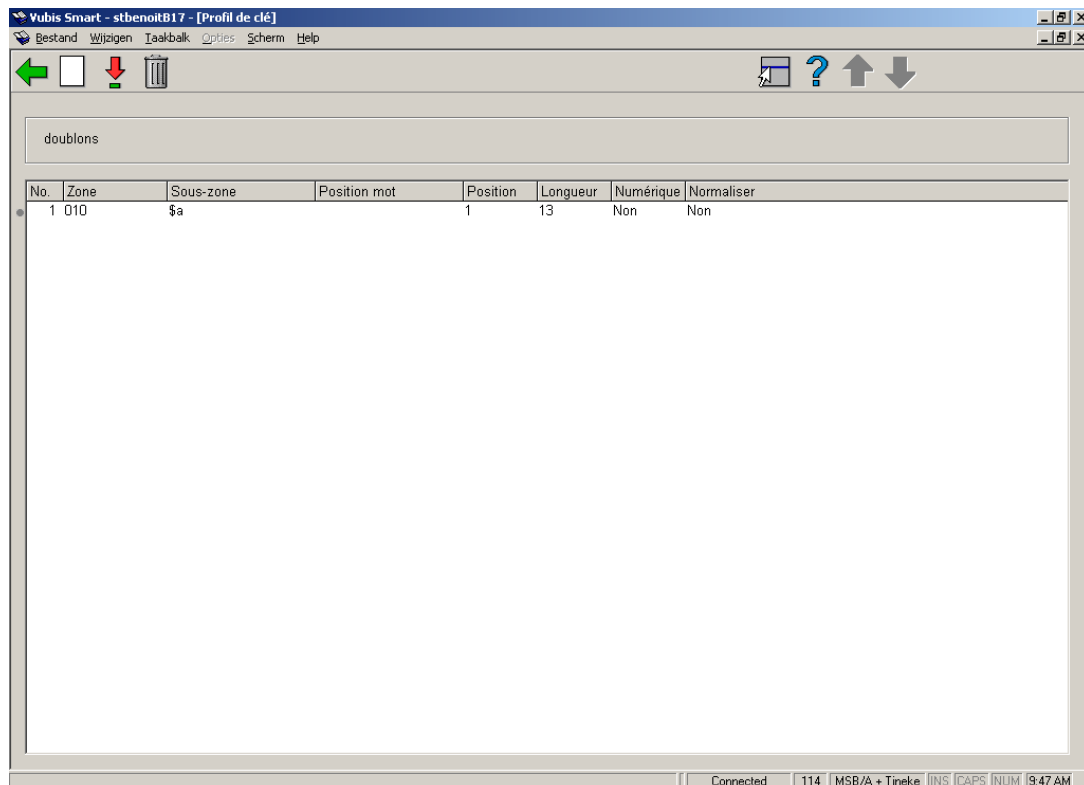
Notes

Le format des enregistrements affichés dans la liste déroulante, dépend des formats définis dans votre système Vubis Smart (Unimarc, Marc/21, etc.). Pour avoir d'autres formats d'enregistrement, consultez votre Administrateur système.

Consulter/modifier propriétés (+): choisissez un profil puis cette option pour modifier les propriétés générales. Voir ci-dessus pour une description de l'écran de saisie.

Supprimer (+): choisissez un profil puis cette option pour le supprimer. Le système demandera une confirmation.

Éléments (+): choisissez un profil puis cette option pour modifier la définition liée. Dans ce cas l'écran suivant s'affichera:



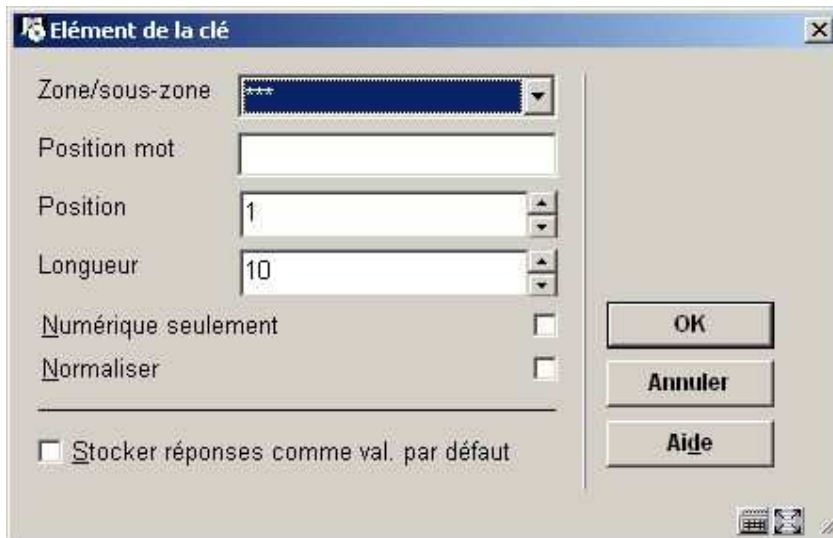
The screenshot shows the Vubis Smart software interface. The window title is "Vubis Smart - stbenoitB17 - [Profil de clé]". The menu bar includes "Bestand", "Wijzigen", "Taakbalk", "Opties", "Scherm", and "Help". The toolbar contains icons for back, forward, search, and other functions. The main area displays a window titled "doublons" containing a table with the following data:

No.	Zone	Sous-zone	Position mot	Position	Longueur	Numérique	Normaliser
1	010	\$a		1	13	Non	Non

The status bar at the bottom indicates "Connected", "114", "MSB/A + Tineke", "JINS", "CAPS", "NUM", and "9:47 AM".

Options sur l'écran

Nouveau: Après la sélection de cette option, un écran de saisie s'affichera :



Zones sur l'écran

Zone/sous-zone : choisir la zone ou la sous-zone sur laquelle doit porter la sélection

Position mot : position du mot dans la sous-zone.

Position : position de départ.

Longueur : longueur des données

Numérique seulement : restriction de la recherche sur des données numériques

Normaliser : utilisé pour normaliser des données (conversion en majuscule et suppression des caractères non alphanumériques à l'exception des espaces entre les données).

Consulter/modifier propriétés (+): choisissez une ligne puis cette option pour modifier les propriétés générales. Voir ci-dessus pour une description de l'écran de saisie.

Supprimer (+): choisissez une ligne et puis cette option pour la supprimer.

Notes

Plus le nombre de critères définis est élevé, plus les chances d'identifier des doublons sont faibles ; la conséquence de ceci est que de véritables doublons risquent de ne pas être détectés. Plus le nombre de critères définis est faible, plus les chances d'identifier des faux doublons sont élevées; la conséquence de ceci est qu'un grand nombre de descriptions risquent à tort d'être présentées par le système comme semblables. Il est donc recommandé de définir un nombre « raisonnable » de critères. Ainsi, vous ne pouvez par exemple pas indiquer que le seul critère pris en compte est que la première lettre du titre et la première lettre du premier auteur doivent être identiques. Une telle sélection est si large que le résultat de la consolidation serait vide de sens.

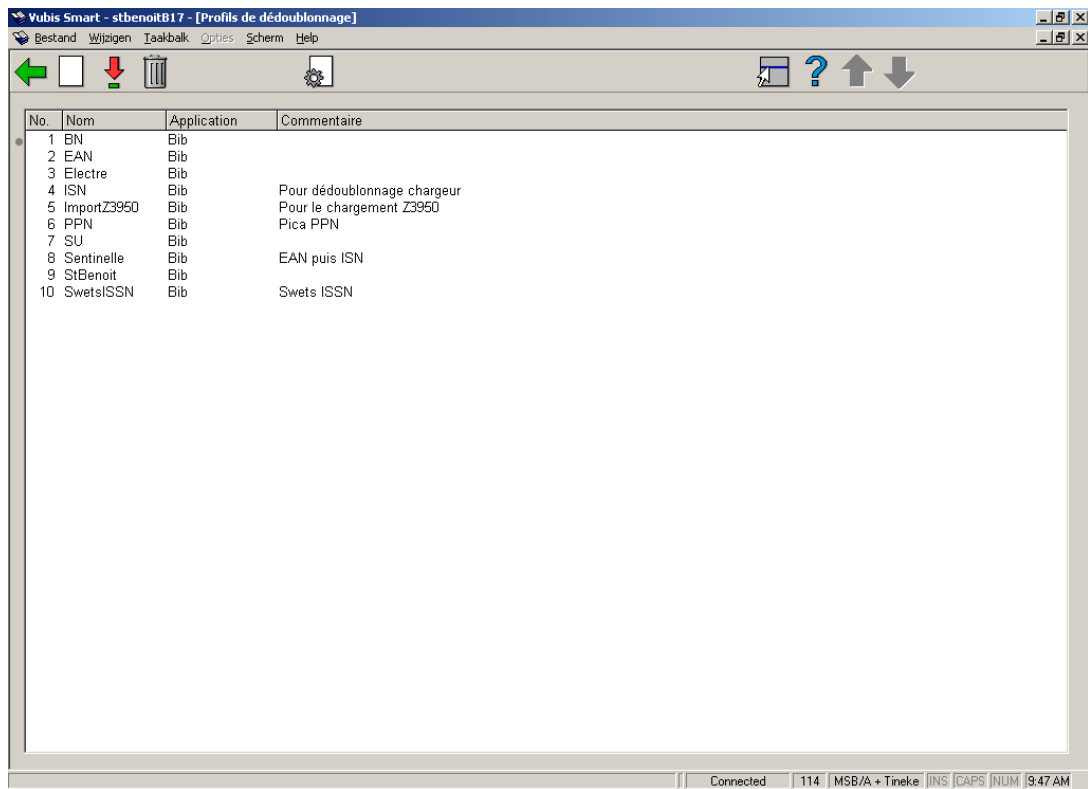
114.3 Profils de dédoublonnage

Un profil de dédoublonnage détecte les notices identiques et détermine comment le système doit réagir face à des notices qui contiennent des clés identiques. Les règles enregistrées dans les profils déterminent ce qui doit être fait avec les notices. Ces règles peuvent ressembler à :

- « Si plusieurs notices avec une clé identique ont été trouvées, les fusionner. »
- « Si une notice est importée et qu'aucune clé identique n'a été trouvée, créer une nouvelle notice »

En résumé : une règle détermine ce qui doit être fait dans le cas où il y a zéro, une ou plusieurs correspondances avec une clé.

Après la sélection de cette option, un écran sommaire s'affichera :



Ces profils d'importation peuvent être employés ailleurs, par exemple pour importer des titres.

Options sur l'écran

Nouveau: Après la sélection de cette option, un écran de saisie s'affichera :



Zones sur l'écran

Nom : introduisez le nom de votre nouveau profil de dédoublement.

Application: cliquez sur la flèche vers le bas pour sélectionner :

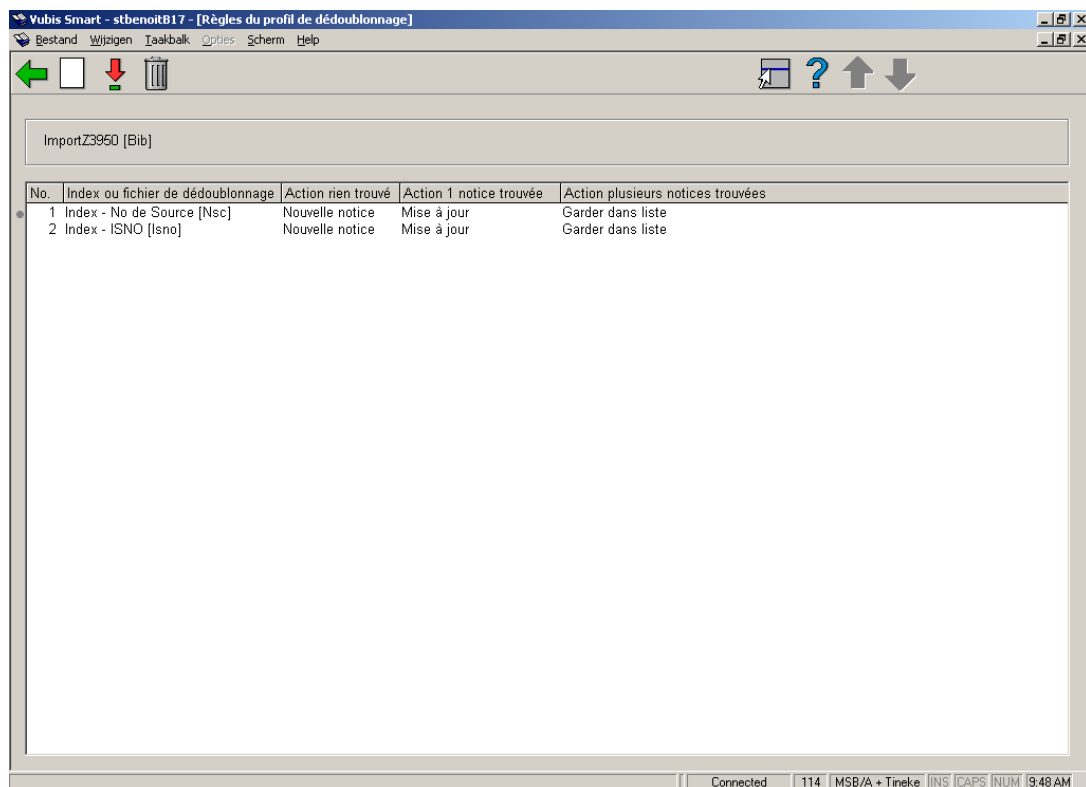
- **Bib** - pour dédoubler des notices bibliographiques.
- **Auth** - pour dédoubler des enregistrements autorisés.

Commentaire : introduisez un texte décrivant votre profil de dédoublement.

Supprimer (+): choisissez une ligne puis cette option pour la supprimer.

Propriétés: choisissez une ligne puis cette option pour modifier les propriétés générales. Voir ci-dessus pour une description de l'écran de saisie.

Consulter/modifier règles(+): choisissez une ligne puis cette option pour modifier les règles. Dans ce cas l'écran suivant s'affichera :

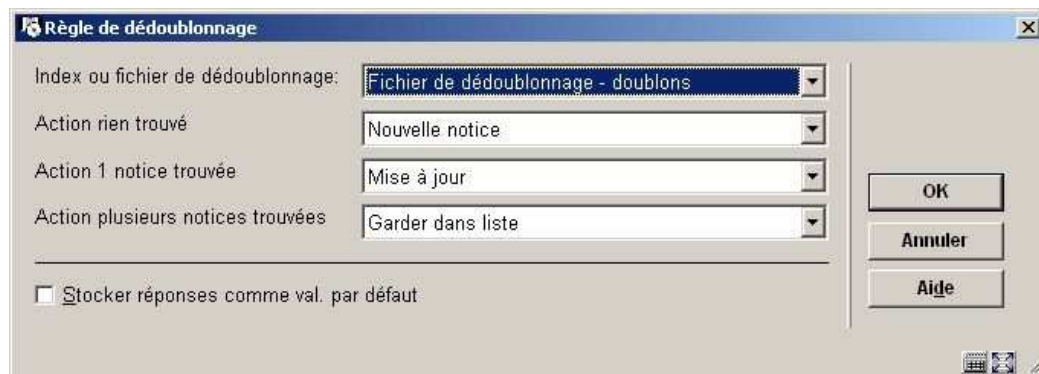


L' ISNO est un critère qui permet de retrouver les notices doublons dans cet exemple. Puisque l'index existe, la clé n'est absolument pas nécessaire. Nous choisissons l'index ISBN, l'action 0 signifie : nouvelle notice, action 1 signifie : mise à jour de la notice, action plusieurs notices trouvées : mettre les notices dans un panier.

La conséquence est que quand l'ISNO n'est pas trouvé dans l'index, une nouvelle notice est ajoutée. Dans le cas où il y aurait une correspondance dans la base, la nouvelle notice serait fusionnée avec la notice existante. Dans le cas où plusieurs notices seraient trouvées, le système ne peut déterminer avec quelle notice la notice entrante doit être fusionnée. Elle sera donc sauvegardée dans un panier. Dans le cas où vous choisissez 'mise à jour' également pour plusieurs notices trouvées, le système fusionnera la notice entrante avec la première notice trouvée dans l'index ISNO.

Options sur l'écran

Nouveau: Après la sélection de cette option, un écran de saisie s'affichera :



Zones sur l'écran

Index ou fichier de dédoublement : un index ou un fichier de dédoublement peut servir de critère de recherche de doublons ; en sélectionner un dans le menu déroulant.

Action rien trouvé : considérer la notice entrante comme nouvelle notice quand le système n'a pas trouvé de notice locale correspondante

Action 1 notice trouvé : considérer la notice entrante comme une mise à jour de la notice locale quand le système a trouvé une notice locale correspondante

Action plusieurs notices trouvées : considérer la notice entrante comme à garder dans un fichier de travail lorsque la notice entrante correspond à plusieurs notices locales.

Supprimer (+): choisissez une ligne puis cette option pour la supprimer.

Consulter/modifier propriétés (+): choisissez une ligne et puis cette option pour modifier les propriétés générales. Voir ci-dessus pour une description de l'écran de saisie.

114.4 Fichiers de dédoublement Un fichier de dédoublement est un pseudo index qui contient des clés uniques pour les notices. L'écran sommaire montre tous les fichiers créés (via l'AFO 114, option 4). A partir d'ici, vous pouvez les supprimer. La consultation/modification des contenus est effectuée dans l'AFO 115.

La constitution d'un fichier de dédoublement nécessite qu'il soit d'abord créé sur le système avec ses paramètres, puis lancé pour aller chercher les doublons sur la base en fonction des critères désirés.

Après la sélection de cette option, un écran sommaire s'affichera :

No.	Nom	Commentaire	Profil de clé	Application	Base de données	Liste	Statut	Clés	Notices
1	doublons		doublons	Bib	St Benoit		Fini	861/1/0	1706/3/0

Colonnes sur l'écran

Nom: Nom du fichier

Commentaire: Description explicative

Profil de clé: Profil utilisé pour ce fichier

Application: Bib - pour dédoublonner des notices bibliographiques ou Auth - pour dédoublonner des notices autorités

Base de données: Nom valide de base de données

Liste: Nom du panier traité (videz quand la base de données complète a été traitée)

Statut: Le statut actuel (fini ou Création)

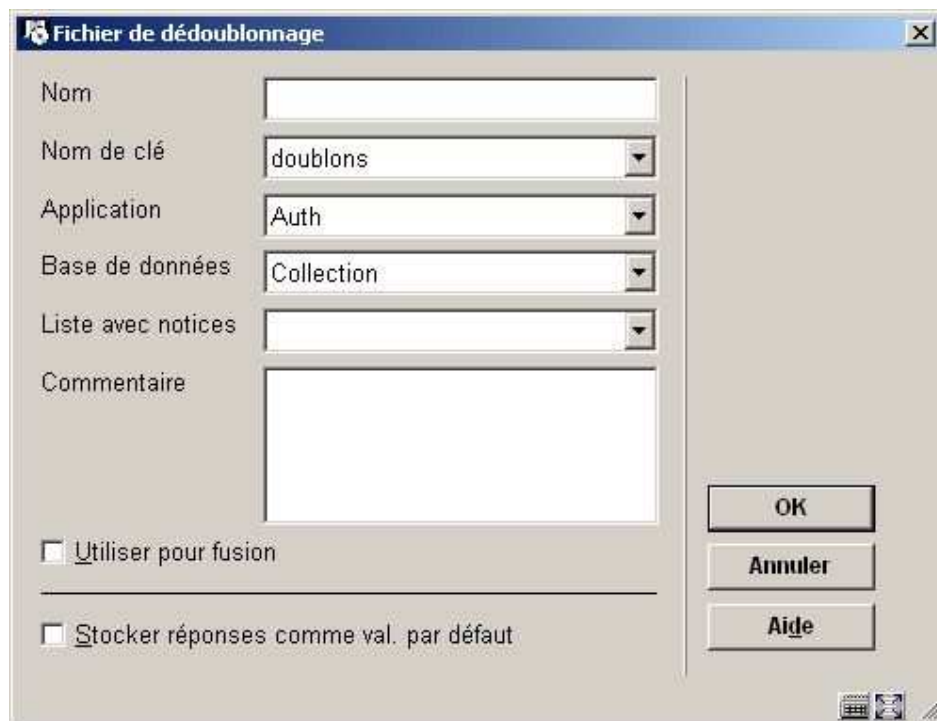
Clés: Nombre de clés (nombre total/fusionné/supprimé)

Notices: Nombre de notices (nombre total/fusionné/supprimé)

Quand l'option « Utiliser pour fusion » a été sélectionnée lors de la création du fichier de dédoublonnage, le fichier peut être utilisé dans l'AFO 115 pour fusionner les notices. C'est uniquement dans ce cas que les notices dupliquées seront sauvegardées dans un panier.

Options sur l'écran

Nouveau: Après la sélection de cette option, un écran de saisie s'affichera :



Zones sur l'écran

Nom : introduisez le nom (de moins de 25 caractères) de votre nouveau fichier de dédoublonnage.

Nom de clé: récupérer la clé précédemment créée

Application: : « Bib » pour les notices bibliographiques et « Auth » pour les notices d'autorité

Base de données : sélectionner le nom de votre base de données

Liste avec notices : il est possible de fusionner deux types de fichiers : l'intégralité du catalogue bibliographique ou un panier. La fusion de notices reprises dans un panier est prévue pour pouvoir sélectionner/fusionner des notices d'une partie du catalogue (par exemple uniquement les CD ou les rapports). Ne pas remplir cette zone si la sélection doit porter sur l'ensemble du catalogue. Saisir le nom du panier si vous souhaitez fusionner des notices reprises dans celui-ci.

Commentaire : introduisez un texte décrivant votre profil de dédoublonnage.

Utiliser pour fusion : cocher cette case si vous souhaitez utiliser un profil de fusion une fois les doublons détectés.

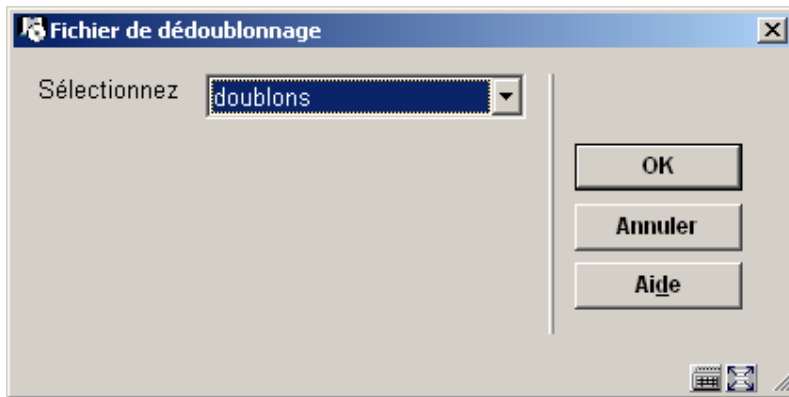
Consulter/modifier propriétés (+): choisissez une ligne et puis cette option pour modifier les propriétés générales. Voir ci-dessus pour une description de l'écran de saisie.

Supprimer (+): choisissez une ligne et puis cette option pour la supprimer.

Rafraîchir: choisissez cette option pour rafraîchir l'écran. Tous les nouveaux fichiers supplémentaires lancés par un autre utilisateur s'afficheront également au même moment.

114.5 Lancement du fichier de dédoublonnage

La définition d'un fichier de dédoublonnage est faite dans AFO 114 sous l'option 2 puis créée avec l'option 4. Après la sélection de cette option, un écran de saisie s'affichera :



Dans la zone **Sélectionnez**, cliquez sur la flèche vers le bas pour sélectionner le fichier de dédoublement que vous désirez recréer.

Attention !

Cette opération nécessite énormément de ressource et de temps, il est donc vivement recommandé de la lancer en fin de journée et de prévoir de la place disque suffisante.

Consulter le fichier de fusion à partir de l'AFO 115 – Option Fichier de fusion.

Après avoir (re)créé le fichier vous pouvez trouver ce fichier sous **Fichiers de dédoublement** dans l'AFO 114

- **Document control - Change History**

Version	Date	Change description	Author
1.0	June 2007	Creation	
1.1	August 2007	Modifications	