

# 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.

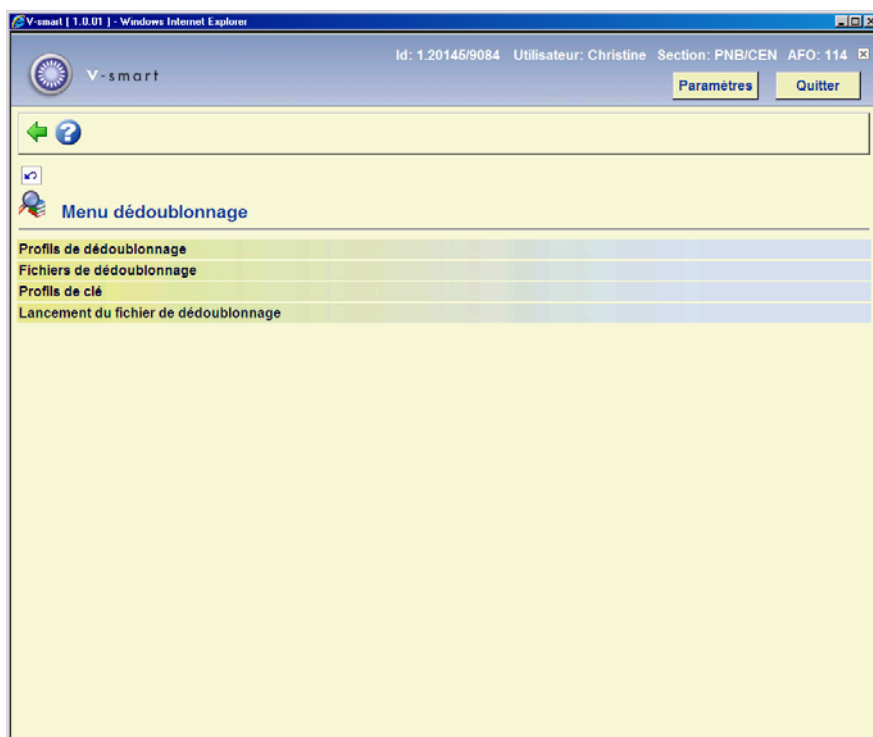
Les AFOs concernées

- 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 »

Formatted: Pattern: Clear

Deleted: ¶  
¶  
¶

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



Deleted: ¶  
¶

Deleted: ¶  
¶

Pour effectuer une recherche de doublons :

Formatted: Pattern: Clear

- 1. Premièrement, il faut créer un profil de clé (option 3)
- 2. Deuxièmement, il faut créer un fichier de dédoublonnage (option 2)
- 3. Pour finir, il faut lancer le fichier de dédoublonnage (option 4).

Deleted: ¶  
¶

Deleted: ¶  
¶  
¶

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 :



The screenshot shows the V-smart application interface. At the top, it displays the user information: 'Id: 1.20145/9084 Utilisateur: Christine Section: PNBI/CEN AFO: 114'. Below this, there are navigation icons and a table titled 'Profils de clé'. The table has three columns: 'No.', 'Nom', and 'Commentaire'. It lists 15 entries, each with a checkbox in the 'No.' column.

| No.                         | Nom               | Commentaire   |
|-----------------------------|-------------------|---|
| <input type="checkbox"/> 1  | 200\$a            | Lets check the 200\$a   |
| <input type="checkbox"/> 2  | 245\$a            | look at 245\$a  |
| <input type="checkbox"/> 3  | ISBN              |   |
| <input type="checkbox"/> 4  | Marc2102\$a       | Marc2102\$a   |
| <input type="checkbox"/> 5  | SmartAuth         |   |
| <input type="checkbox"/> 6  | UMSubject         |   |
| <input type="checkbox"/> 7  | UniMarcAuth       | Subjects  |
| <input type="checkbox"/> 8  | authorized fields | Added this one in VSMART  |
| <input type="checkbox"/> 9  | controle          |   |
| <input type="checkbox"/> 10 | debs              | 010\$a is defined here  |
| <input type="checkbox"/> 11 | frakeydef1        | key definition for Smart  |
| <input type="checkbox"/> 12 | frakeydef2        | 2nd key definition for Smart  |
| <input type="checkbox"/> 13 | multiple keys     | Test LOTS of characters in the key definition for Matching profiles |
| <input type="checkbox"/> 14 | philtest10        | 010 without subfields   |
| <input type="checkbox"/> 15 | smartauthority    | HDR16409  |

Deleted: ¶  
¶

Deleted: ¶  
¶

Deleted: ¶  
¶

## Options sur l'écran

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



The screenshot shows the 'Profil de clé' configuration screen. It features a form with the following elements:

- Nom:** A text input field.
- Format:** A dropdown menu currently set to 'AuthorityList'.
- Commentaire:** A large text area for notes.
- Buttons:** 'OK', 'Annuler', and 'Aide' buttons are located on the right side.
- Checkbox:** A checkbox labeled 'Stocker réponses comme val. par défaut' is located at the bottom.

Deleted: ¶  
¶

Deleted: ¶  
¶

## 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 V-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.

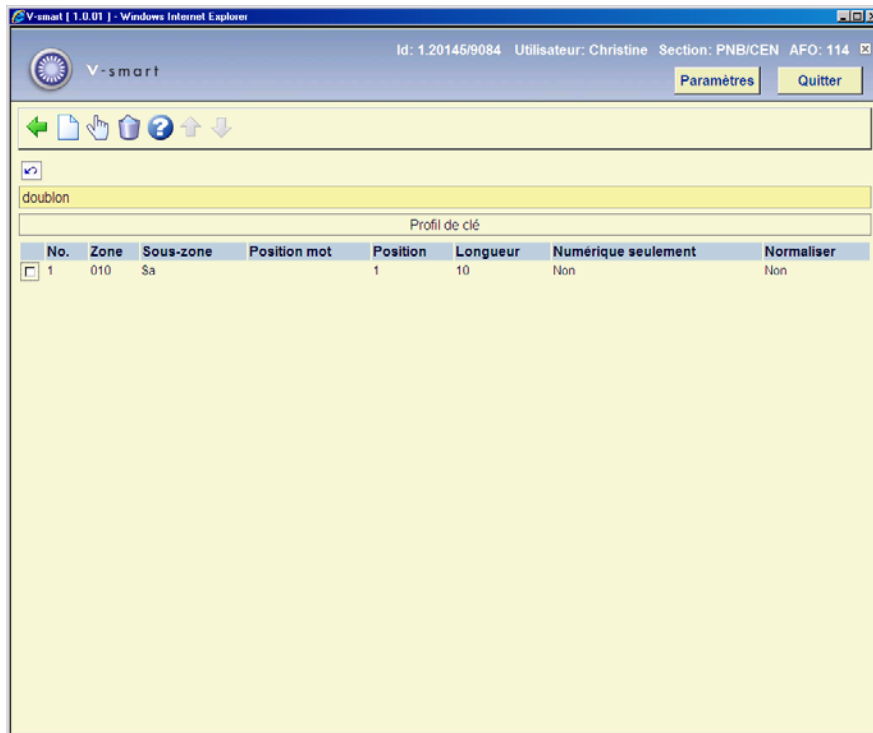
**Supprimer (+)**: choisissez un profil puis cette option pour le supprimer. Le système demandera une confirmation.

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

Deleted: ¶  
¶

Deleted: ¶  
¶

Deleted: ¶  
¶

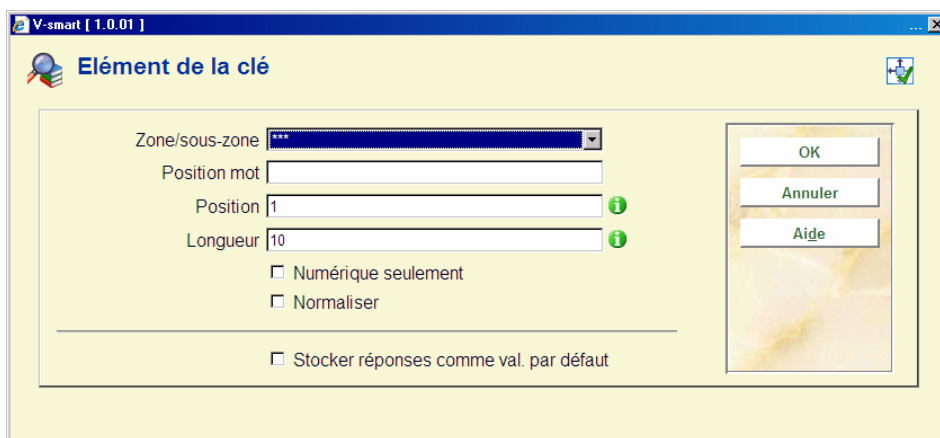


Deleted: ¶  
¶

### Options sur l'écran

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

Deleted: ¶  
¶



### 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).

Deleted: ¶  
¶  
¶

**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.

Deleted: ¶  
¶

**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.

Deleted: ¶  
¶

Deleted: ¶  
¶

## 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 :

Formatted: Pattern: Clear

- « 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.

Deleted: ¶  
¶

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

Deleted: ¶  
¶

The screenshot shows a web browser window titled 'V-smart [ 1.0.01 ] - Windows Internet Explorer'. The page header includes 'Id: 1.20145/9084 Utilisateur: Christine Section: PNB/CEN AFO: 114' and buttons for 'Paramètres' and 'Quitter'. Below the header is a navigation bar with icons for back, forward, home, search, and refresh. The main content area is titled 'Profils de dédoublement' and contains a table with 15 rows. Each row has a checkbox, a number, a name, an application type, and a comment.

| No.                      | Nom | Application     | Commentaire  |
|--------------------------|-----|-----------------|--|
| <input type="checkbox"/> | 1   | Authority Match | Auth HDR16409  |
| <input type="checkbox"/> | 2   | ISBN\$a         | Bib Match on ISBB \$A                                    |
| <input type="checkbox"/> | 3   | ISSN M21        | Bib <Font color='red'>RED COMMENTS</font>                |
| <input type="checkbox"/> | 4   | M21Tword        | Bib  |
| <input type="checkbox"/> | 5   | M21recorddelte  | Bib Check multiple indexes for deletion                  |
| <input type="checkbox"/> | 6   | Marc21Authors   | Bib  |
| <input type="checkbox"/> | 7   | Marc21ISBN      | Bib Marc21ISBN   |
| <input type="checkbox"/> | 8   | PPN             | Bib Pica PPN   |
| <input type="checkbox"/> | 9   | SmartAuthor     | Bib  |
| <input type="checkbox"/> | 10  | SwetsISSN       | Bib Swets ISSN   |
| <input type="checkbox"/> | 11  | UMSubjects      | Auth   |
| <input type="checkbox"/> | 12  | deb             | Bib deb's 1st profile - Index4 Smart ISBN and a savelist |
| <input type="checkbox"/> | 13  | fratest         | Bib test 14/11   |
| <input type="checkbox"/> | 14  | testisbn10      | Bib  |
| <input type="checkbox"/> | 15  | title           | Bib smart title (200 \$a)                                |

Deleted: ¶  
¶

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

### Options sur l'écran

Deleted: ¶  
¶

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



Deleted: ¶  
¶

### 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 autorités.

**Commentaire** : introduisez un texte décrivant votre profil de dédoublement.

Formatted: Pattern: Clear

Deleted: ¶  
¶

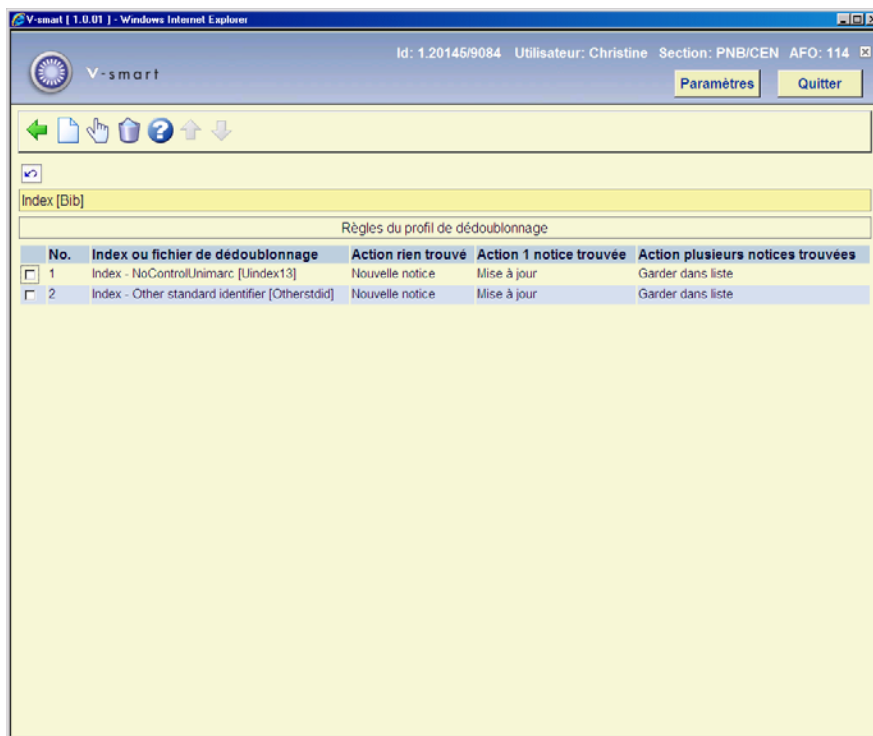
Deleted: ¶  
¶

**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 :





Deleted: ¶  
¶

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.

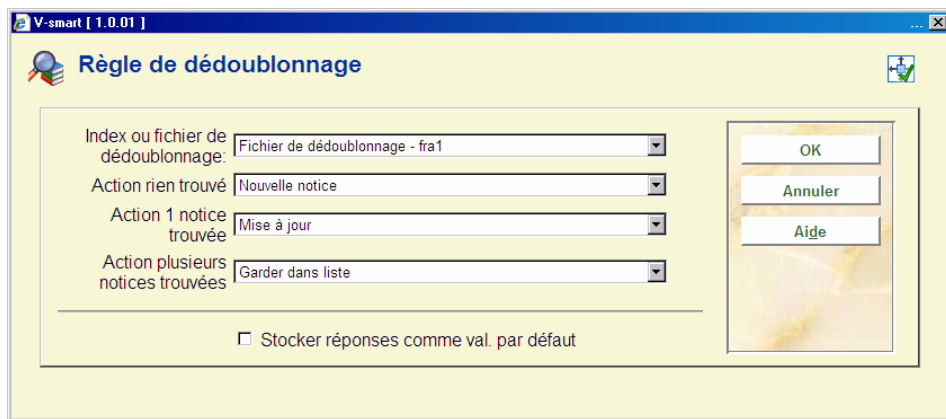
La conséquence est que quand l'ISNO n'est pas trouvé dans l'index, une nouvelle notice est ajoutée. Dans le cas où il y aurait une correspondance dans la base, la nouvelle notice serait fusionnée avec la notice existante. Dans le cas où plusieurs notices seraient trouvées, le système ne peut déterminer 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.

Deleted: ¶  
¶

### Options sur l'écran

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

Deleted: ¶  
¶



Deleted: ¶  
¶

## 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.

Deleted: ¶  
¶

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

Deleted: ¶  
¶

**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.

Deleted: ¶  
¶

**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.

Deleted: ¶  
¶

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

V-smart [ 1.0.01 ] - Windows Internet Explorer

Id: 1.20145/9084 Utilisateur: Christine Section: PNB/CEN AFO: 114

Paramètres Quitter

Fichiers de dédoublement

| No.                         | Nom                              | Commentaire  | Profil de clé | Application | Base de données                       | Liste              | Statut | Clés   | Notices |
|-----------------------------|----------------------------------|--|---------------|-------------|---------------------------------------|--------------------|--------|--------|---------|
| <input type="checkbox"/> 1  | HDR16409SmartAuthors             | Smart Authority Authors  | 200\$a        | Auth        | Nom de personne                       |                    | Finis  | 75/2/1 | 192/5/2 |
| <input type="checkbox"/> 2  | ISBNsQA                          | Read from database? Where is the database defined?               | debs          | Bib         | UniMarc/B                             |                    | Finis  | 5/0/0  | 10/0/0  |
| <input type="checkbox"/> 3  | Marc21ISBN                       | ISBNs  | Marc2102\$a   | Bib         | Bibliographic database                | 21RECORDSHARVESTED | Finis  | 0      | 0       |
| <input type="checkbox"/> 4  | New_underscore_by_deb            |  | debs          | Bib         | Vubis Smart database release 2.1.1 B6 |                    | Finis  | 493    | 593     |
| <input type="checkbox"/> 5  | Savelists without the underscore |  | debs          | Bib         | Vubis Smart database release 2.1.1 B6 |                    | Finis  | 489    | 581     |
| <input type="checkbox"/> 6  | Savelists_from_deb_for_Smart_DB  | HDR 14648 We have 187 records in this savelist for SMART format. | debs          | Bib         | Vubis Smart database release 2.1.1 B6 | SCIENCEFICTION     | Finis  | 0      | 0       |
| <input type="checkbox"/> 7  | SmartAuth                        | match 220\$a in Smart Authorities                                | SmartAuth     | Auth        | Mot-clé adultes                       |                    | Finis  | 4/0/0  | 26/0/0  |
| <input type="checkbox"/> 8  | UMSubjects                       |  | UMSubject     | Auth        | Nom commun UNIMARC                    |                    | Finis  | 6/1/0  | 24/2/0  |
| <input type="checkbox"/> 9  | UniMarcAuth                      |  | UniMarcAuth   | Auth        | Nom commun UNIMARC                    |                    | Finis  | 7/0/0  | 26/0/0  |
| <input type="checkbox"/> 10 | debssmartisbn                    | Use for Merge debs smart database here                           | debs          | Bib         | Vubis Smart database release 2.1.1 B6 |                    | Finis  | 93/2/0 | 251/5/0 |
| <input type="checkbox"/> 11 | fra1                             |  | fra1          | Bib         | Vubis Smart database release 2.1.1 B6 | GEAC               | Finis  | 0/0/0  | 0/0/0   |
| <input type="checkbox"/> 12 | fra2                             |  | fra1          | Bib         | Vubis Smart database release 2.1.1 B6 |                    | Finis  | 0      | 0       |
| <input type="checkbox"/> 13 | frances                          |  | frakeydef1    | Bib         | Vubis Smart database release 2.1.1 B6 |                    |        | 0/0/0  | 0/0/0   |
| <input type="checkbox"/> 14 | over256chars                     | Key definition exceeds 256 characters                            | multiple keys | Bib         | Bibliographic database                |                    | Finis  | 5088   | 393     |
| <input type="checkbox"/> 15 | testphilisbn                     |  | philtest10    | Bib         | Vubis Smart database release          | PHIL50             | Finis  | 18     | 18      |

## Colonne sur l'écran

**Nom:** Nom du fichier

**Commentaire:** Description explicative

**Profil de clé:** Profil utilisé pour ce fichier

**Application:** Bib - pour dédoubler des notices bibliographiques ou Auth - pour dédoubler 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 dédoublement, 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.

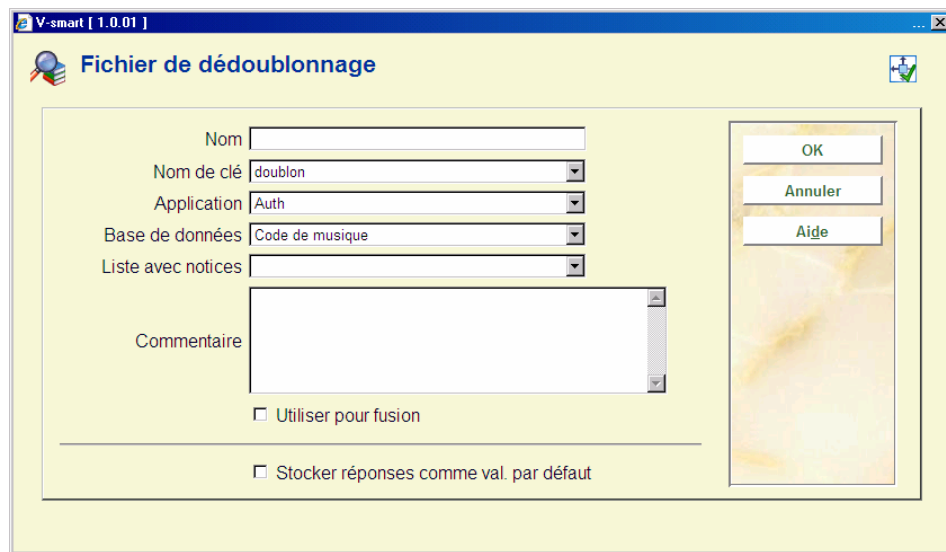
Deleted: ¶  
¶

### Options sur l'écran

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

Deleted: ¶  
¶

Deleted: ¶  
¶



### Zones sur l'écran

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

**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 dédoublement.

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

Deleted: ¶  
¶

**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.

Deleted: ¶  
¶

Deleted: ¶  
¶

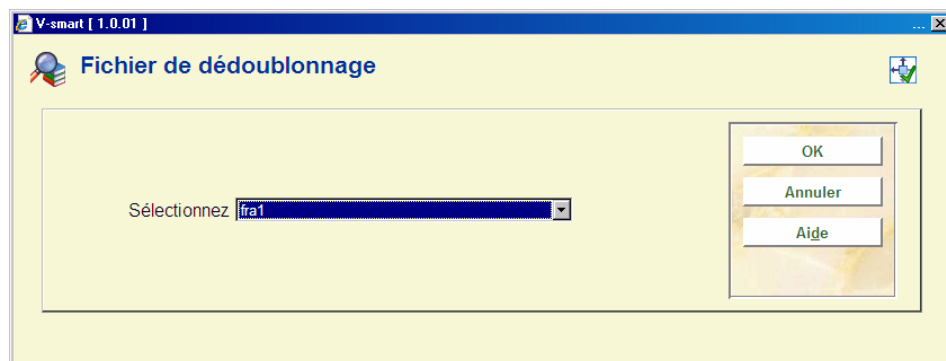
**Supprimer (+)**: choisissez une ligne et puis cette option pour la supprimer.

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

Deleted: ¶  
¶

## 114.5 Lancement du fichier de dédoublement

La définition d'un fichier de dédoublement 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 :



Deleted: ¶  
¶

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

Deleted: ¶  
¶

**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.

Deleted: ¶  
¶

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

Deleted: ¶  
¶

- Document control - Change History

| Version | Date     | Change description | Author |
|---------|----------|--------------------|--------|
| 1.0     | May 2008 | Creation           |        |

Deleted: Christine Lé

Deleted: ¶