AFO 114 – Record matching

114.1 Introduction

Use this AFO to find and process matching records when merging records. Use the Record match menu options to:

- Detect duplicate records when records are loaded or merged into a database file.

- Remove duplicate records from record sets. For example, in the record sets produced by a Z39.50 search.

Record matching allows you to decide if new and duplicate records are to be accepted, rejected, or set aside in a savelist.

The record matching process uses a matching profile. You can define a number of matching profiles. However, the record matching process can use only one matching profile at a time.

The following specialised terms are used:

- Matching Profiles: a set of rules for searching for matching records in a set of records.

- Matching Files: a rule for searching for matching records in a set of records.

- Key definition profiles: a key for determining if records are identical and matching.

- (Re)create Matching files: a type of index containing unique record keys.

After starting this AFO the following submenu is displayed:
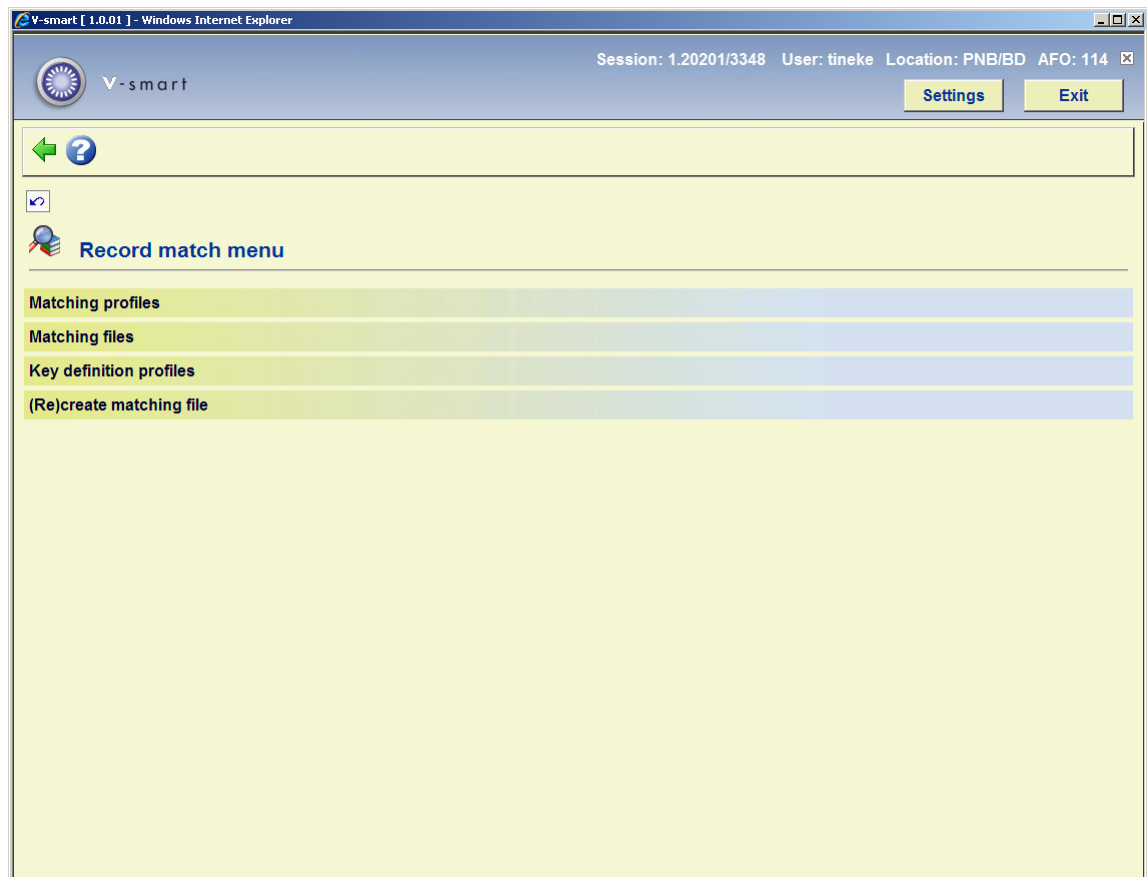
A matching profile is a set of rules for matching records. Use the following steps in order to create a matching profile:

1.   Create one or more key definition profiles (option 3).

2.   Create a matching file (option 2).

3.   Fill the matching file with data (option 4).

4.   Create a matching profile for example for the import of titles, based on the file created.

Steps 1-3 can be used to get an overview of all duplicate titles in the matching file. The file can also be used to compare files to be imported, but you can of course also do this by using a normal index.

Comparing titles is necessary to be able to detect duplicates during importing or merging processes. Comparison is done via indexes or pre/defined keys. The result of a comparison can be zero, one ore multiple duplicates.
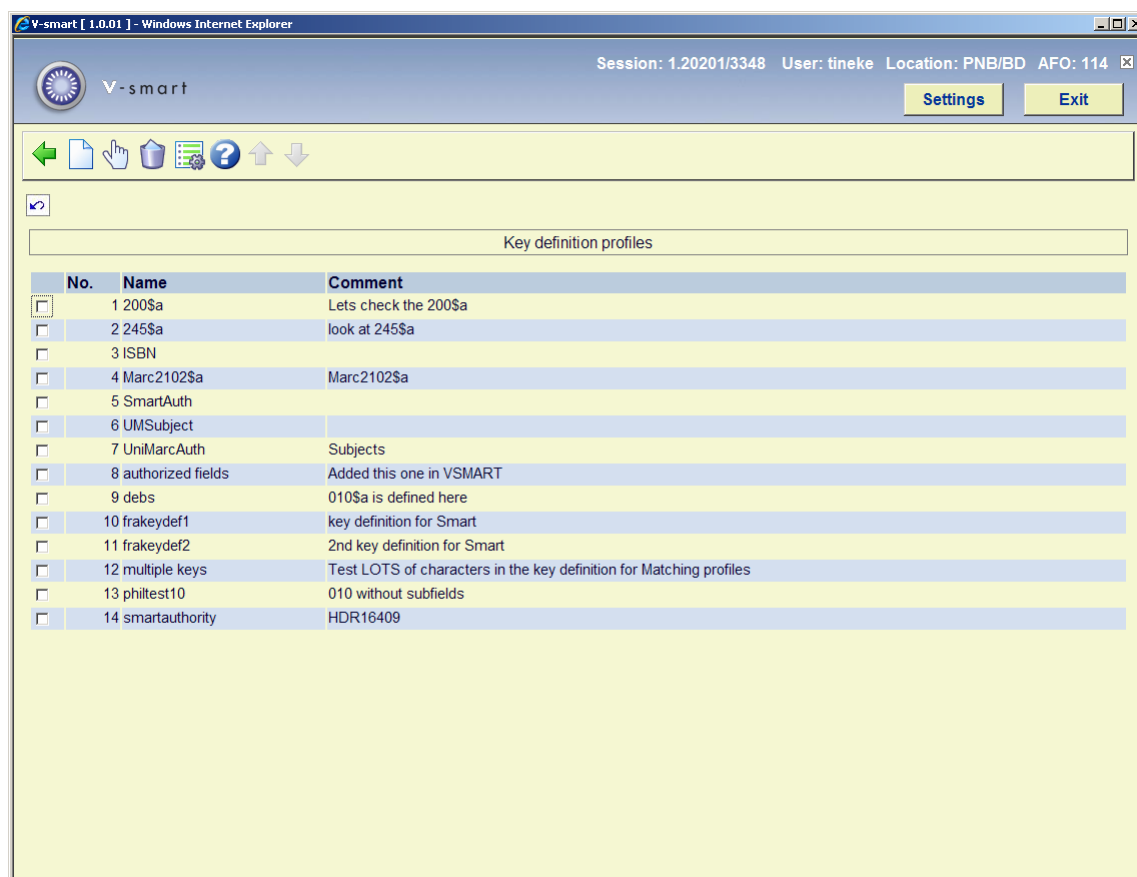
New records as well as duplicates can be accepted, rejected or set aside in a savelist with a special status to be assessed later.

Comparison of records is not only done during imports, but can also be used as a separate action by the user, to assess a file. Apart from that it is also used to compare two different sets of records, for instance after a Z39.50 search. In this case so called matching files are used.

114.2 Key definition profiles

A key consists of elements from a record. This is a string of characters which can look like "financial^jone^elsev^2004" (random example). The elements of the key are derived from the record, in the example: the first word of the title, the first four characters of the author name, the first five characters of the publisher name and the year of publication. Because elements of a key can be derived from repeatable fields)like author name, multiple keys per title are possible. For comparison all available keys are used.

After choosing this option the following overview screen is displayed:

Options on the screen

New item: choose this option to create a new key definition profile.

View-modify item properties (+): select a profile and then this option to modify the general properties.

Delete item (+):select a profile and then this option to delete it. The system will ask for confirmation.

Elements (+):select a profile and then this option to modify the linked definition. In that case the following screen will be displayed:
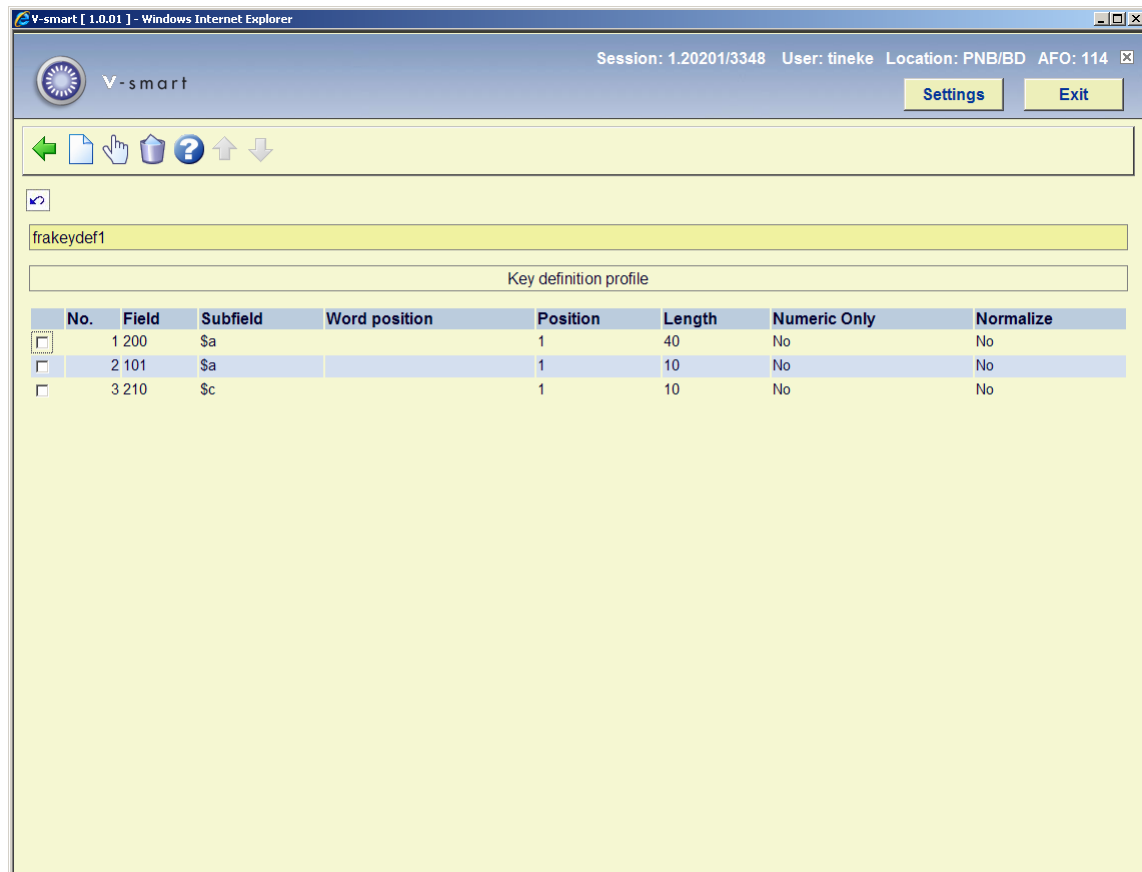

Options on the screen

New item: choose this option to create a new definition.

View-modify item properties (+): select an item and then this option to modify the general properties.

Delete item (+):select an item and then this option to delete it.

In the example below we see: of subfield 200$a (main title) the first 20 characters of the word are taken. This data is normalised (everything to uppercase, punctuation except spaces stripped). Of subfield 700$b (author last name) the first 4 characters of the field are taken and

also normalised. Then the first 5 characters of the publisher name are taken. Then we take the first 4 characters of the publication year, numeric only so the actual data remains (and additions like cop., ed. are ignored).
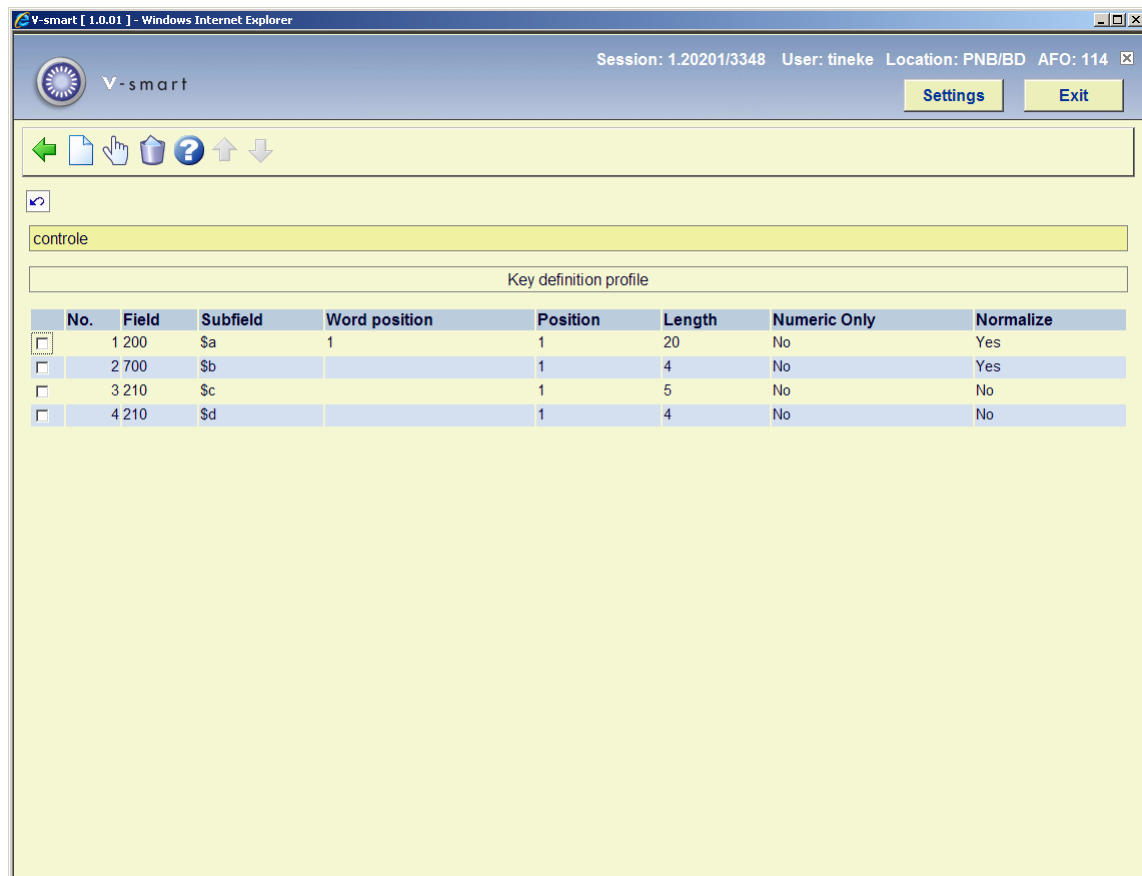
The system supports creation of the keys and saving them, so they can be used again and again, for instance when importing records, searching titles via Z39.50 etc.

But the process can also be done "on the fly" without permanently saving the keys.

114.3 Matching profiles and rules

A matching profile detects identical records and determines how the system must react to records that have identical keys. The rules saved in the profiles determine what must be done with the records. These rules can be something like this:

- "If multiple records with an identical key are found, merge them".

- "If a record is imported and no identical keys are found, create a new record".

To summarise: a rule determines what must be done in case there are zero, one or multiple matches with a key.

The profile screen looks like this:
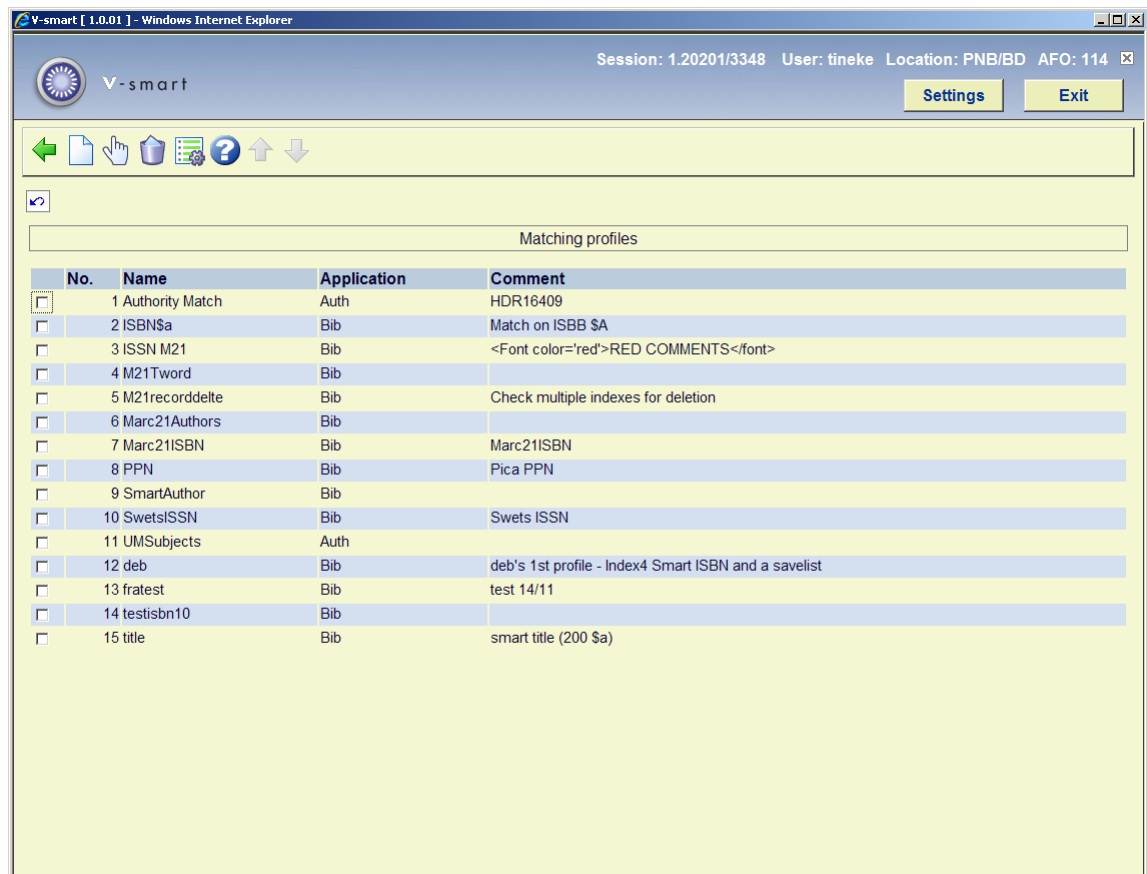
These import profiles can be used elsewhere, for instance for importing titles.

Options on the screen

New item: choose this option to create a new matching profile definition.

Delete item (+):select a profile and then this option to delete it.

Properties (+):select a profile and then this option to modify the general properties.

View/modify item rules (+):select a profile and then this option to modify the linked definition. In that case the following screen will be displayed:
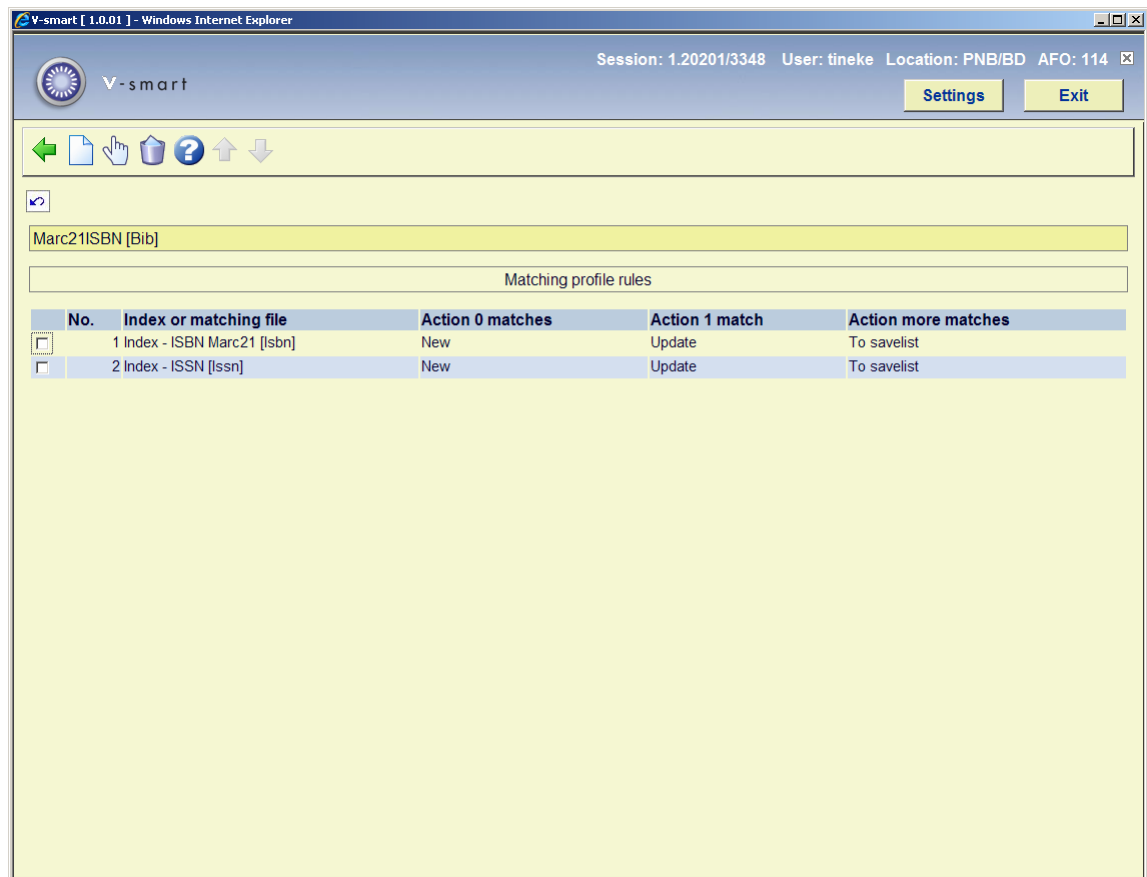
The ISBN is the criterion for finding duplicate records in this example. Because there is an index on ISBN the key is not absolutely necessary. We choose the ISBN index, action 0 matches: new record, action 1 match: update record, action multiple matches: put records in a savelist.

The effect is that when no ISBN is found in the index, a new title record will be created. In case there is one match, the new title will be merged with the existing title. In case of multiple matches, the system cannot determine with which record the incoming record must be merged. It will therefor be put in a savelist. In case you choose 'update record' also for multiple matches, the system will merge the incoming record with the first matching record encountered in the ISBN index.

Options on the screen

New item: choose this option to create a new definition.

View/modify item properties (+): select a file and then this option to modify the general properties.

Delete item (+):select a file and then this option to delete it.

114.4 Matching files

A matching file is a pseudo index with unique keys for the records. The overview screen shows all files created (via AFO 114, option 4). From here you can delete them. Viewing/modifying the contents is done in AFO 115.

The files are shown in a list:

Columns on the screen

Name: Name of the file

Comment: Explanatory description (what the file is for)

Key definition profile: Profile used for this file

Application: Bibliographic (titles) or authorities (e.g. authors)

Database: Valid database name

Savelist: Name of processed savelist (empty when the complete database has been processed)

Status: Ready or Processing

Keys: Number of keys (number/merged/deleted)

Records: Number of records (number/merged/deleted)

When the option "use for merge" has been selected when creating the matching file, the file can be used in AFO 115 to merge records. In that case only duplicate records will be saved in the file.

Options on the screen

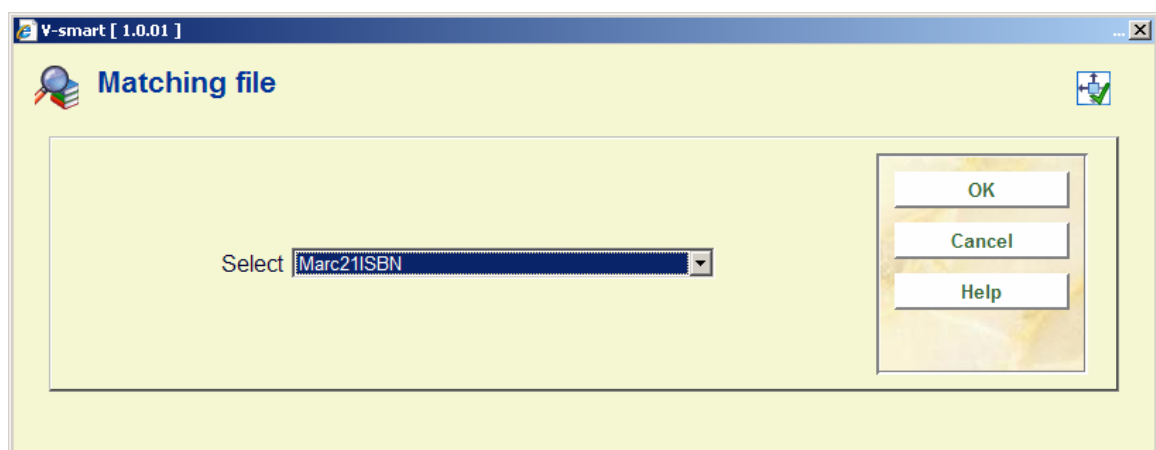**New item**: choose this option to create a new file.

View/modify item properties (+): select a line and then this option to modify the general properties.

Delete item (+):select a line and then this option to delete it..

Refresh: select this option to refresh the screen. Any new files added by another user in the mean time will be shown as well now.

114.5 Recreate matching file

A definition for a matching file is made in AFO 114 under option 2 en then built with option 4. After choosing this option an input form will be displayed:

After (re)creating you can find the file under "Matching files" in AFO 114.

- **Document control - Change History**

Version	Date	Change description	Author
1.0	April 2008	creation	