

# AFO 114 - Vergelijken van documenten

## 114.1 Inleiding

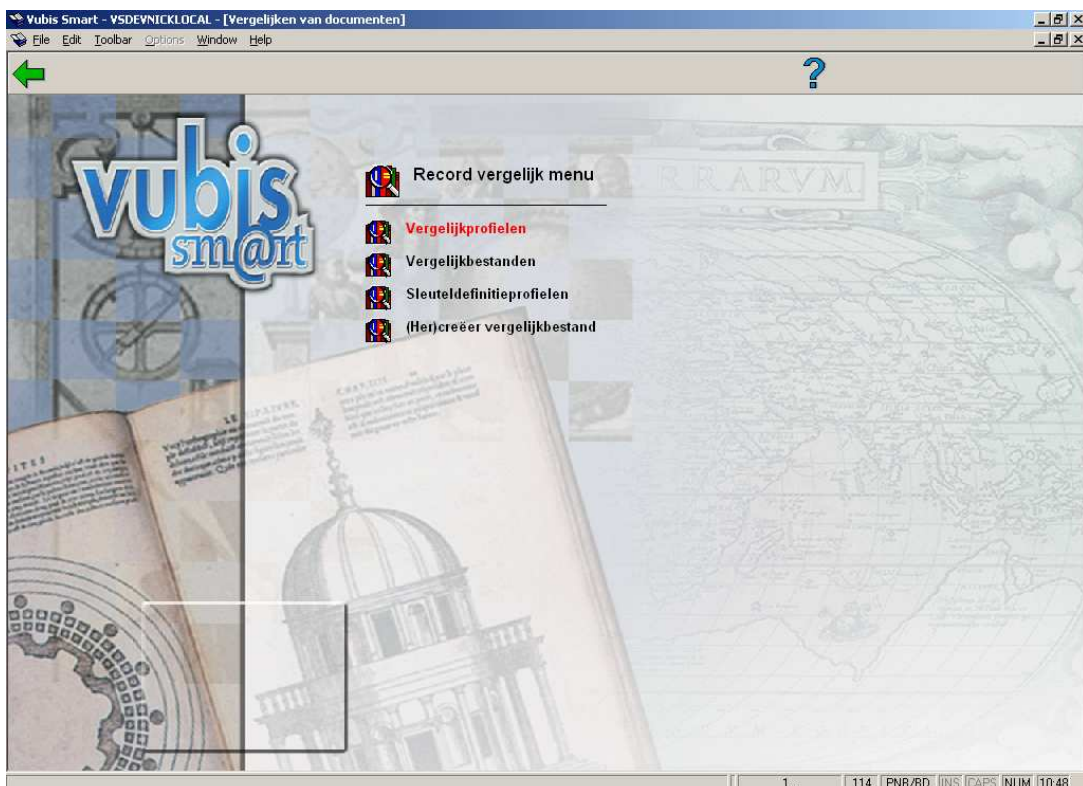
Met behulp van AFO 114 kunt u titelbeschrijvingen vergelijken als voorbereiding op het samenvoegen van gelijke records. Gebruik deze AFO voor:

- Het opsporen van duplicaten wanneer er records geladen of samengevoegd zijn in de database.
- Het verwijderen van duplicaten.

Het vergelijken van records stelt u in staat te beslissen of duplicaatrecords moeten worden geaccepteerd, verwijderd of bewaard in een savelist.

Het vergelijkingsproces maakt gebruik van bepaalde profielen. U kunt deze zelf definiëren, zo veel als nodig. Het vergelijkingsproces kan echter maar 1 profiel tegelijk gebruiken.

Nadat u deze AFO heeft gestart verschijnt onderstaand submenu:



De meest uitgebreide werkwijze is als volgt:

1. Maak een of meerdere sleuteldefinities aan (optie 3).
2. Bouw een vergelijkbestand (optie 4).
3. Bekijk/bewerk het vergelijkbestand (optie 2).
4. Maak een vergelijkprofiel voor bv de import van titels op basis van het gemaakte bestand.

Stappen 1 t/m 3 kunt u gebruiken om een overzicht te krijgen van al bestaande dubbele titels in het bestand. Het bestand kan ook gebruikt worden om te importeren bestanden te vergelijken, maar hiervoor kan ook gewoon een index worden gebruikt.

Nieuwe PICA records hebben bijvoorbeeld een uniek PPN nummer. Indien er al een index op dit nummer in het bestand aanwezig is, volstaat het om alleen een vergelijkprofiel "PPN" te maken gebaseerd op de PPN index in plaats van een vergelijkbestand. Hieronder gaan we verder in op de principes van het vergelijken.

Het vergelijken van titels is noodzakelijk om dubbele titels te kunnen detecteren tijdens een import of samenvoegproces. Het vergelijken gebeurt door middel van indexen of voorgedefinieerde sleutels. Het resultaat van een vergelijking kan nul, een of meerdere duplicaten zijn.

Zowel nieuwe als dubbele records kunnen worden geaccepteerd, afgewezen of opgeslagen in een bewaarlijst (met een speciale status) om later te beoordelen wat er mee moet gebeuren.

Het vergelijken gebeurt door middel van een vergelijkingsprofiel. Meerdere van deze profielen kunnen worden gedefinieerd, maar bij de import van titels kan er maar een tegelijk worden gebruikt.

Het vergelijken van records wordt niet alleen tijdens de import gebruikt, maar kan ook los door de gebruiker worden aangeroepen om een bestand te beoordelen. Daarnaast wordt het ook gebruikt om twee verschillende recordsets te kunnen vergelijken, bijvoorbeeld na een z39.50 zoekactie. In dit geval worden zogenaamde vergelijkbestanden gebruikt.

Vubis gebruikt de volgende principes en terminologie:

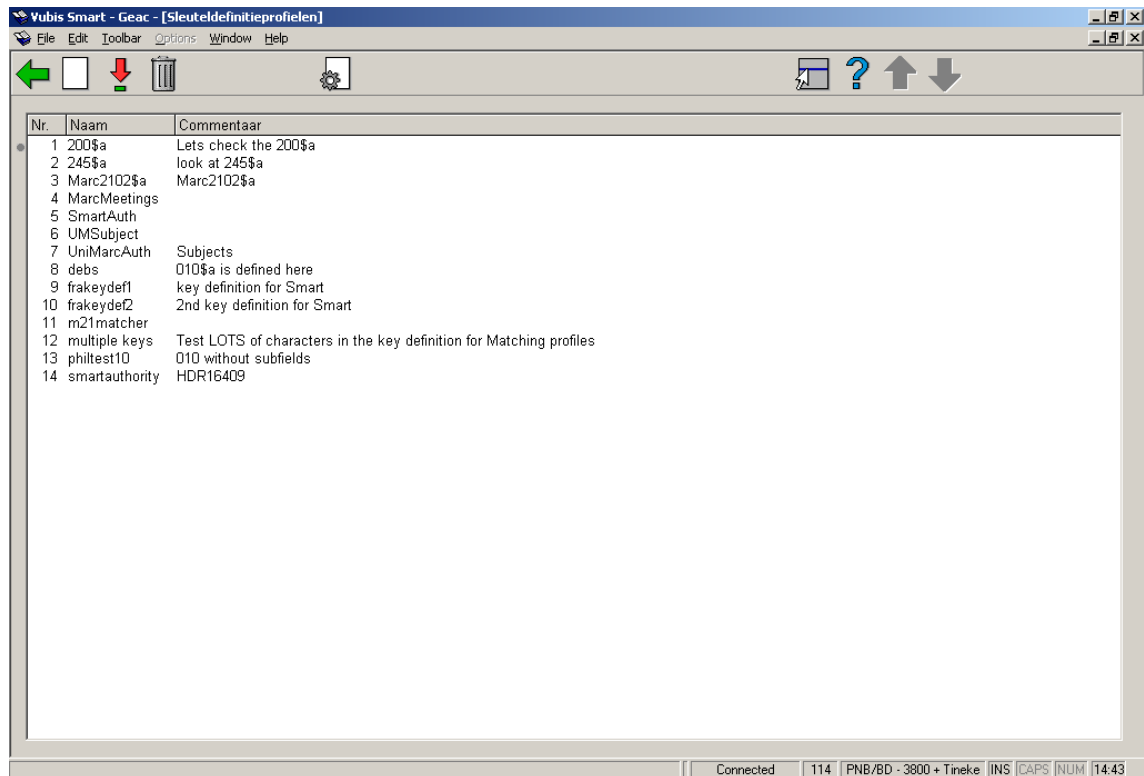
- (Vergelijkings)sleutels: een sleutel wordt gebruikt door het systeem om mogelijk identieke records te kunnen opsporen.

- Vergelijkingsprofielen: een vergelijkingsprofiel is een verzameling regels die aangeeft hoe het systeem te werk moet gaan wanneer er gezocht wordt naar dubbele records.
- Vergelijkingsregels: een regel geeft aan hoe het systeem moet reageren wanneer er dubbele records gevonden worden in een vergelijking.

## 114.2 Sleuteldefinitie

Een sleutel bestaat uit elementen die afkomstig zijn uit een record. Het is een verzameling van karakters die eruit kan zien als “financieel^veen^elsev^1999” (willekeurig voorbeeld). De elementen van de sleutel zijn afgeleid uit het record, in dit geval bijvoorbeeld het eerste woord van de titel, de eerste vier letters van de auteur, de eerste vijf letters van de uitgever en het jaar van uitgave. Omdat de elementen van een sleutel kunnen worden afgeleid van herhaalbare velden (bijvoorbeeld auteurs), zijn er per titel meerdere sleutels mogelijk. Voor het vergelijken worden alle beschikbare sleutels gebruikt.

Nadat u deze optie heeft gekozen wordt onderstaand overzichtsscherm getoond:



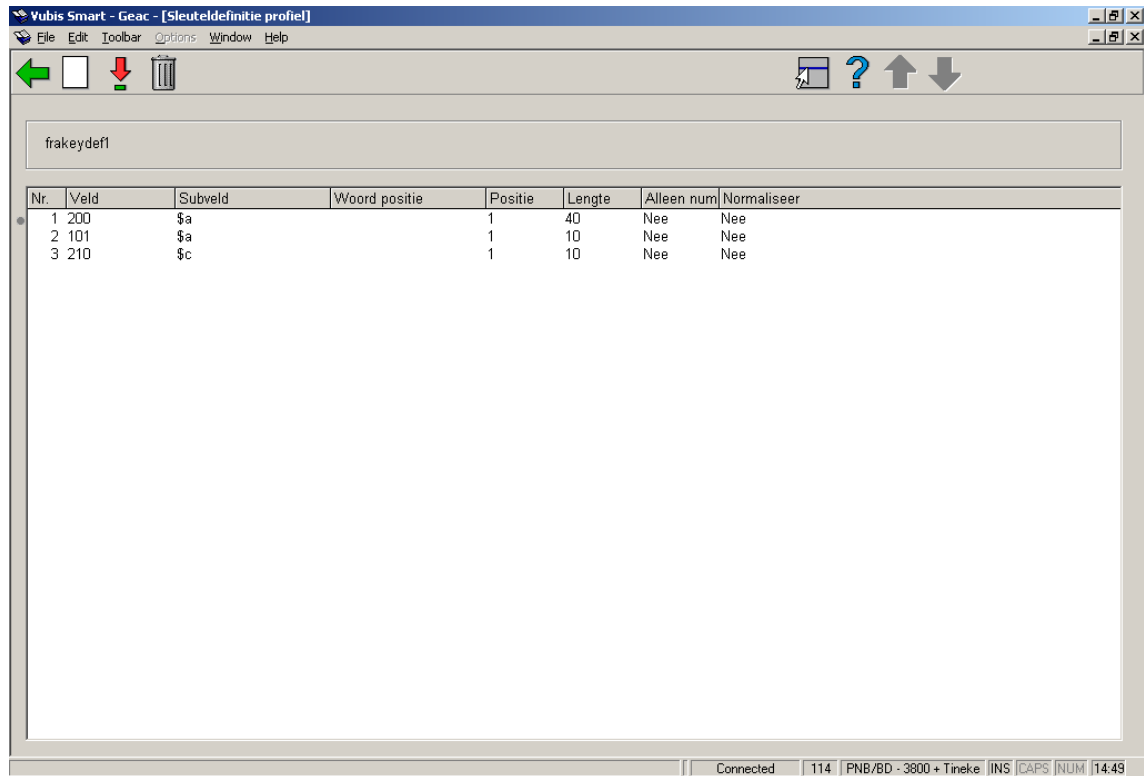
### Opties op het scherm

**Nieuw item:** kies deze optie om een nieuw sleuteldefinitieprofiel aan te maken.

**Bekijken/wijzigen item eigenschappen (+):** selecteer een profiel en dan deze optie om de algemene eigenschappen te wijzigen.

**Schrap item (+):** selecteer een profiel en dan deze optie om het profiel te verwijderen. Het systeem zal om een bevestiging vragen

**Elementen (+):** selecteer een profiel en dan deze optie om de bijbehorende definitie te wijzigen. In dat geval verschijnt onderstaand scherm:



The screenshot shows a window titled 'Vubis Smart - Gear - [Sleuteldefinitie profiel]'. The window contains a toolbar with icons for back, forward, and search. Below the toolbar is a text field containing 'frakeydef1'. The main area of the window displays a table with the following data:

Nr.	Veld	Subveld	Woord positie	Positie	Lengte	Alleen num	Normaliseer
1	200	\$a		1	40	Nee	Nee
2	101	\$a		1	10	Nee	Nee
3	210	\$c		1	10	Nee	Nee

The status bar at the bottom of the window shows 'Connected', '114', 'PNB/BD - 3600 + Tineke', 'INS', 'CAPS', 'NUM', and '14:49'.

## Opties van het scherm

**Nieuw item:** kies deze optie om een nieuwe definitie toe te voegen.

**Bekijken/wijzigen item eigenschappen (+):** selecteer een regel en dan deze optie om de algemene eigenschappen te wijzigen.

**Schrap item (+):** selecteer een regel en dan deze optie om deze te verwijderen.

In het voorbeeld hieronder zien we dit uitgewerkt; van veld 200/\$a (hoofdtitel) worden de eerste 20 posities van het eerste woord genomen. Deze gegevens worden genormaliseerd (alles naar hoofdletter en interpunctie eruit behalve spaties). Van 700/\$b (achternaam auteur) nemen we de eerste vier karakters van het gehele veld, ook weer genormaliseerd. Daarna de

eerste 5 karakters van de uitgever. Als laatste nemen we uit het veld jaar van uitgave de eerste vier karakters, alleen numeriek zodat het jaartal overblijft (cop. of ed. in het veld wordt dan genegeerd).

Nr.	Veld	Subveld	Woord positie	Positie	Lengte	Alleen numeriek	Normaliseer
1	200	\$a	1	1	20	Nee	Ja
2	700	\$b		1	4	Nee	Ja
3	210	\$c		1	5	Nee	Nee
4	210	\$d		1	4	Ja	Nee

Het systeem heeft de mogelijkheid om de sleutels te maken en op te slaan, zodat deze vervolgens bij meerdere gelegenheden kunnen worden gebruikt, bijvoorbeeld bij een importeeractie, zoeken op titels via Z39.50, etc.

Daarnaast kan dit proces ook “on the fly” gebeuren zonder dat de sleutels permanent worden opgeslagen.

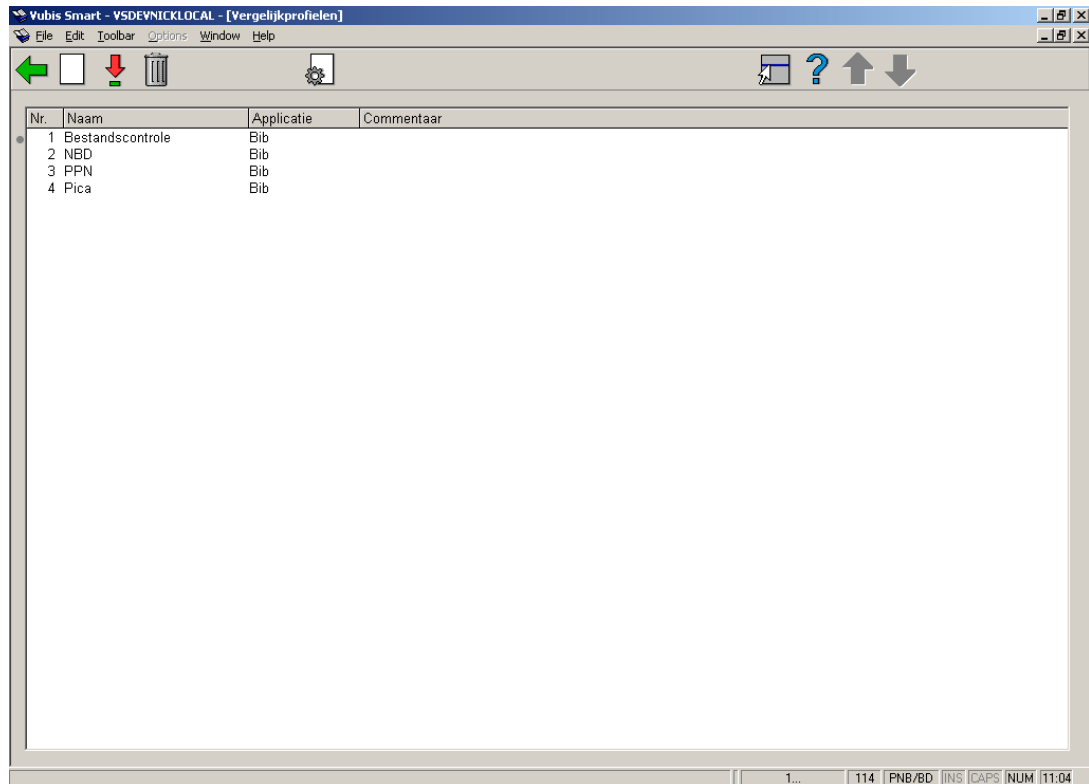
## 114.3 Vergelijkingsprofielen en regels

Een vergelijkingsprofiel detecteert identieke records en bepaalt hoe het systeem moet reageren op deze titels die een identieke sleutel hebben. De regels die opgeslagen liggen in het profielen bepalen wat er met de records moet gebeuren. Deze regels kunnen er bijvoorbeeld als volgt uitzien:

- “Als meerdere records met dezelfde sleutel worden gevonden, voeg ze dan samen”.
- “Als een record wordt geïmporteerd en er zijn geen overeenkomende sleutels, maak dan een nieuwe titel aan”.

Samengevat bepaalt een regel wat er moet gebeuren in het geval van nul, een of meerdere matches met een sleutel.

Het profielscherm ziet er zo uit:



Deze beschikbare profielen kunnen elders ingezet worden, bijvoorbeeld in de titelimport.

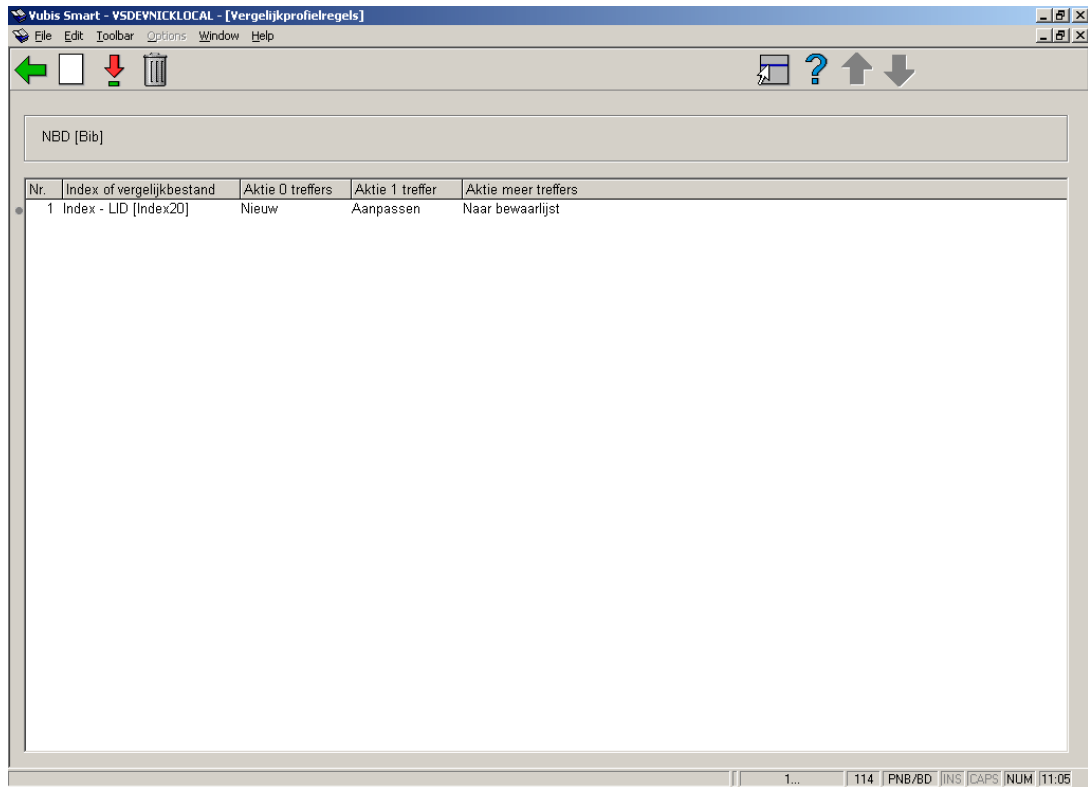
### Opties op het scherm

**Nieuw item:** kies deze optie om een nieuw vergelijkingsprofiel aan te maken.

**Schrap item (+):** selecteer een profiel en dan deze optie om het profiel te verwijderen.

**Elementen (+):** selecteer een profiel en dan deze optie om de algemene eigenschappen te wijzigen.

**Bekijken/wijzigen item eigenschappen (+):** selecteer een profiel en dan deze optie om de bijbehorende definitie te wijzigen. In dat geval verschijnt onderstaand scherm:



Het LID nummer is in dit geval het criterium om tot dubbele records te komen. Er is een index op LID nummer, dus een sleutel is niet echt noodzakelijk. We kiezen voor de LID index; actie 0 treffers="nieuw"; 1 treffer="aanpassen"; meerdere treffers="naar bewaarlijst".

Dit heeft als effect dat wanneer een LID nummer niet wordt gevonden in de index er een nieuwe titel wordt aangemaakt. Wordt er een match gevonden, dan wordt de nieuwe titel met de oude samengevoegd. Worden er meerdere bestaande titels gevonden met hetzelfde LID nummer dan is voor het systeem niet te bepalen met welk record moet worden samengevoegd, dus gaat de nieuwe titel naar een bewaarlijst voor latere beoordeling. U kunt als actie bij meer treffers ook "aanpassen" activeren, de nieuwe titel wordt dan samengevoegd met de eerste overeenkomende titel die in de LID index wordt gevonden.

### Opties van het scherm

**Nieuw item:** kies deze optie om een nieuwe definitie toe te voegen.

**Bekijken/wijzigen item eigenschappen (+):** selecteer een regel en dan deze optie om de algemene eigenschappen te wijzigen.

**Schrap item (+):** selecteer een regel en dan deze optie om deze te verwijderen.

**114.4 Vergelijkbestanden** Een vergelijkbestand is een pseudo-index met daarin unieke sleutels voor records. Het overzichtsscherm laat alle aangemaakte bestanden (AFO 114, optie 4) zien. Van hieruit kunnen ze worden verwijderd. Het bekijken/bewerken van de inhoud van de lijst gebeurt in AFO 115.

De bestanden worden weergegeven in een lijstpresentatie:

Nr.	Naam	Commentaar	Sleuteldefinitie prof	Applicatie	Bestand	Bewaarlijst	Status	Sleutels	Records
1	HDR16409Smar	Smart Authority	200\$a	Auth	Personele auteur		Klaar	75/2/1	192/5/2
2	ISBNsQA	Read from databa	debs	Bib			Klaar	5/0/0	10/0/0
3	Marc21ISBN	ISBNs	Marc2102\$a	Bib			Klaar	29	27
4	New_underscor		debs	Bib	Vubis Smart data		Klaar	493	593
5	Savellists without		debs	Bib	Vubis Smart data		Klaar	489	581
6	Savellists_from_	HDR 14648 We h	debs	Bib	Vubis Smart data SCIENCEFICTIO		Klaar	304	349
7	SmartAuth	match 220/\$a in	SmartAuth	Auth	Trefwoord volwas		Klaar	4/0/0	26/0/0
8	UMSubjects		UMSubject	Auth	Nom commun		Klaar	6/1/0	24/2/0
9	UniMarcAuth		UniMarcAuth	Auth	Nom commun		Klaar	7/0/0	26/0/0
10	debsmartisbn	Use for Merge de	debs	Bib	Vubis Smart data		Klaar	62/1/0	156/3/0
11	fra1		fra1	Bib	Vubis Smart data GEAC		Klaar	0/0/0	0/0/0
12	fra2		fra1	Bib	Vubis Smart data		Klaar	0	0
13	frances		frakeydefl	Bib	Vubis Smart data		Klaar	0/0/0	0/0/0
14	over256chars	Key definition ex	multiple keys	Bib			Klaar	5088	393
15	testphilisbn		philtest10	Bib	Vubis Smart data PHIL50		Klaar	18	18
16	titletest	add a comment h	200\$a	Bib	Vubis Smart data		Klaar	1118	1168

### Kolommen op het scherm

**Naam:** Naam van het bestand

**Commentaar:** Commentaar mbt het bestand (waar is het bestand voor)

**Sleuteldefinitieprofiel:** Profiel dat gebruikt is voor dit bestand

**Applicatie:** Bibliografisch (titels) of authorities (bv auteurs)

**Bestand:** Geldige bestandsnaam



**Bewaarlijst:** Naam van verwerkte savelist (leeg wanneer de gehele database is verwerkt)

**Status:** Klaar of in opbouw

**Sleutels:** Aantal sleutels (aantal/samengevoegd/verwijderd)

**Records:** Aantal records (aantal/samengevoegd/verwijderd)

Wanneer bij het aanmaken van een vergelijkbestand de optie “gebruik voor samenvoegen” is aangevinkt kan het bestand in AFO 115 worden gebruikt om records samen te voegen. Dan zullen alleen dubbele records in het bestand worden opgeslagen.

### Opties van het scherm

**Nieuw item:** kies deze optie om een nieuw bestand toe te voegen.

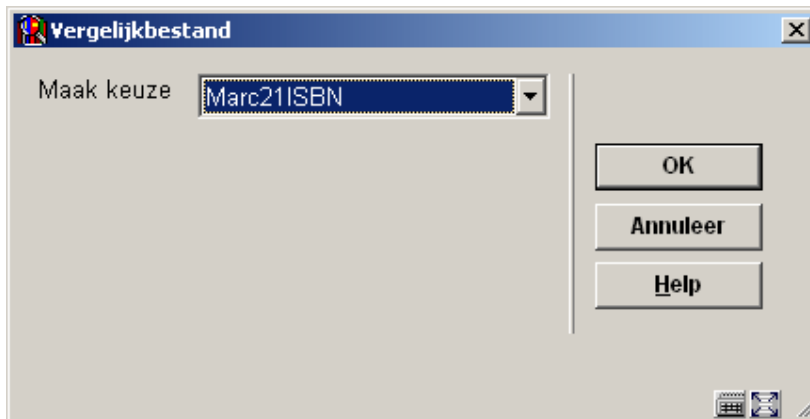
**Bekijken/wijzigen item eigenschappen (+):** selecteer een bestand en dan deze optie om de algemene eigenschappen te wijzigen.

**Schrap item (+):** selecteer een bestand en dan deze optie om deze te verwijderen.

**Scherm verversen:** kies deze optie om de display te verversen. Alle bestanden die in de tussentijd door een andere gebruiker zijn aangemaakt worden nu ook getoond.

## 114.5 Vergelijkbestand herbouwen

Een definitie voor een vergelijkbestand wordt gemaakt in AFO 114 optie 2, en vervolgens gebouwd met optie 4. Nadat u deze optie heeft gekozen verschijnt een invulscherm:



Na het (her)bouwen kunt u het bestand vinden onder "vergelijkbestanden" in AFO 114.

- **Document control - Change History**

<b>Version</b>	<b>Date</b>	<b>Change description</b>	<b>Author</b>
<b>1.0</b>	<b>unknown</b>	<b>Creation</b>	
<b>2.0</b>	<b>May 2006</b>	<b>Various revisions</b> <b>Delivered as part of build 17 set</b>	