

# DH – Data harvester

## 1 Inleiding

Vubis ondersteunt een automatische 'harvester' voor bibliografische records, op basis van het Z39.50 harvesting protocol.

Deze 'harvester' automatiseert de workflow voor catalogiseren op basis van records uit andere databases, waarbij Z39.50 catalogiseerprofielen gebruikt worden om Z39.50 bronnen te doorzoeken en vervolgens gegevens in uw eigen catalogusrecords te voegen.

Hieronder volgt een beschrijving van de benodigde stappen om deze functionaliteit te kunnen gaan gebruiken.

## 2 Setup

Er moet een vergelijkprofiel gedefinieerd worden in AFO 114 om aan te geven welke veld(en) er gebruikt moeten worden om te bepalen of records hetzelfde zijn.

Er moet een samenvoegprofiel gedefinieerd worden in AFO 115 om aan te geven wat er moet gebeuren met velden wanneer u records samenvoegt die 'hetzelfde' zijn gebaseerd op het vergelijkprofiel.

Optioneel kan er een conversieprofiel gedefinieerd worden in AFO 134, in het geval het formaat van de inkomende records afwijkt van het formaat in uw eigen database.

Er wordt een standaard profiel gedefinieerd in AFO 133 waarin het Z39.50 type van het inkomende record is vastgelegd, alsmede benodigde conversies, match & merge criteria, doeldatabase en sjabloon, e.d. Dit wordt gekoppeld aan de eerder gedefinieerde vergelijk- en samenvoegbestanden (en optioneel aan het conversieprofiel).

Dit importprofiel wordt vervolgens gespecificeerd in de definitie van de Z39.50 Database in AFO 651. Er wordt een Z39.50 profiel gedefinieerd voor de partners waarvan gekopieerd wordt, alsmede een catalogiseerprofiel om de originele recordsleutel te koppelen aan de bron waaruit gekopieerd wordt. Zie AFO 651 - Z39.50 Parameters - Target Profielen - Catalogiseerprofielen voor details.

Er wordt een selectieprofiel gemaakt in AFO 141 om een lijst van bibliografische records op te halen. Er zijn twee velden om het invoerscherm voor het maken van een selectieprofiel, die alleen geldig zijn voor bibliografische selecties: 'Update type' en 'Update profiel'.

Raadpleeg de online help van bovengenoemde AFO's voor gedetailleerde informatie over iedere stap.

### 3 Workflow

Het selectieprofiel kan worden uitgevoerd op een voorgedefinieerde bewaarlijst of gebruikt worden om een nieuwe bewaarlijst aan te maken. Het systeem doorloopt de resulterende bewaarlijst en verwerkt elk record als volgt:

1. Er wordt een Z39.50 zoekactie gestart die gebruik maakt van de zoekcriteria als gedefinieerd in het opgegeven catalogiseerprofiel. Dit kan bijvoorbeeld EAN, ISSN, ISBN of andere specifieke zoekleutel zijn.
2. De zoekvraag resulteert in een van drie mogelijke opties:
  - er komt geen record terug – er wordt een melding weggeschreven in het rapportagebestand
  - er komt één record terug – het bibliografische record wordt bijgewerkt met de standaard laad en wijzig opties
  - er komt meer dan één record terug – er wordt een melding weggeschreven in het rapportagebestand

Het rapportagebestand kan worden geraadpleegd via de standaardopties in AFO 642 in standaard txt formaat.

#### Voorbeeld:

We hebben een aantal basisrecords, opgeslagen in een bewaarlijst:




Document	Acties	Bewerken
***	0000nam##2200000#a#4500	
008	091124s enk 000   eng d	
020 ## \$a	0140508791	
245 10 \$a	A halloween mask	
	\$b Virginia Mueller	



De details zijn:

3. Definieer een conversieprofiel in AFO 134 indien nodig, om vast te leggen in welke velden van de lokale database informatie uit inkomende records moet worden opgeslagen.
4. Definieer een Importprofiel in AFO 133. Dit bepaalt karakterset, formaat van binnenkomende records, te gebruiken conversieprofiel, te gebruiken bewaarlijsten tijdens het laadproces en de koppelingen naar vergelijkprofiel, samenvoegprofiel, doeldatabase, conversieprofiel e.d. Aangezien het te gebruiken Harvesting protocol Z39.50 is, dient het import formaat correct ISO2709 formaat te zijn.

 **Invoer/uitvoerprofiel**


Algemeen   **Formaat**   Invoer   Bewaarlijsten   Exemplaren


Vergelijkprofiel


Samenvoegprofiel

Database

Default template

Maximaal aantal fouten voor record  

Maximaal aantal programma fouten voor import  

Maximal aantal programma fouten voor laden  

Record wijzig-status wordt

Record verwijder-status wordt

Als gegevens gelijk aan afgekeurde vorm

Antwoorden opslaan als default

OK  
Annuleer  
Help

5. Maak in AFO 651 - Z39.50 een database en database groep voor de te benaderen Z39.50 server. Geef bij de Data Source definitie het importprofiel op dat is gedefinieerd voor Harvesting (match and merge) in AFO 133.

V-smart [ 2.0.02 ] -- Webpage Dialog

### Database details

Name: ForHarvesting

Wording for the user: useForHarvesting

Internet Address: 163.1.13.199

Port nr: 210

Bestandsnaam (Z39.50 DatabaseName): ADVANCE

Comments:

In Use?

Vraag om login en wachtwoord?

Groep (voor login):

Login naam:

Wachtwoord:

Herhaal wachtwoord:

Kan bladeren (Z39.50 Scan)?

Ondersteunt 'Named Result Sets'?

Preferred Record Syntax: MARC21

Importprofiel (voor conversie): HarvestM21WithOverwrite (Bib)

OK

Annuleer

Help

http://nlbawqavubis/sm100dev/ClientBody.csp?ClientId=1.20003 Local intranet

6. Maak in AFO 651 - Z39.50 - Target Profiel een target of target groep die gebruikt zal worden voor de zoekvraag.

V-smart [ 2.0.02 ] -- Webpage Dialog

### Modify a Target Profile

Naam:   
 Wording:   
 Commentaar:

Database Groups:
 

IVubis-LocalM21	ForHarvesting
AUS-MacQuarie	
BIGGROUP	
CDN-Montréal	
MIT	
NLC-BNC	
NYU	

Vergelijkprofiel:

Database(s) for Matching Profile:
 

Algemene materiaal aanduiding (Auth/1001)	Bibliographic database(Bib/6)
Auteurs - Collectivités et Congrès(Auth/278)	
Auteurs - Noms de personnes (Auth/277)	
Background database 1 for	

Use database for record characteristics:

First Search: No retrieve when hits >=  ⓘ  
 First Search: Maximum to retrieve=  ⓘ  
 First Search: Per Present Request=  ⓘ  
 Time out after ... seconds  ⓘ

OID voor gemeenschappelijk:  ⓘ

<http://nlbawqavubis/sm100dev/ClientBody.csp?ClientId=1.20003> Local intranet

7. Maak in AFO 651 - Targets - Catalogiseerprofiel een profiel voor wijzigen of toevoegen

V-smart [ 2.0.02 ] -- Webpage Dialog

### Modify a Target Profile

Naam:   
 Wording:   
 Commentaar:

Target Profile:

Access Point:

Define Search Key using 1 or more field/subfield combinations:

Strip punctuation?

<http://nlbawqavubis/sm100dev/ClientBody.csp?ClientId=1.20003> Local intranet

Selecteer de target profiel groep voor de Z39.50 zoekgroep. Access point is het Use Attribute dat gestuurd zal worden naar de target Z39.50 zoekgroep.

Search key bepaalt de velden(en)/subveld(en) uit het bronrecord die gebruikt moeten worden om de data te identificeren die als zoek sleutel naar de target host gestuurd moet worden.

Er kunnen meerdere velden(en)/subveld(en) opgegeven worden, gescheiden door een komma. (bijv. 020/\$a,022/\$a voor Marc21 ISBN en ISSN veld).

8. In AFO 141 - Selecteren wordt een nieuw selectieprofiel gemaakt

The screenshot shows a web-based dialog box titled "Eigenschappen van selectie profiel 'HarvestForMarc21ISBN' van applicatie 'Bib'". The dialog contains several input fields and a list of buttons. The "Profiel naam" field is filled with "HarvestForMarc21ISBN". The "Commentaar" field contains "Selection profile for 'Harvester'". The "Bestand waarin geselecteerd wordt" dropdown is set to "Format as defined by the MARC21 comitte". The "Updatetype" dropdown is set to "Update bibliografische records". The "Updateprofiel" dropdown is set to "m21ISBN". There is a checkbox for "Antwoorden opslaan als default" which is currently unchecked. On the right side, there are three buttons: "OK", "Annuleer", and "Help". The window title bar indicates "V-smart [ 2.0.02 ] -- Webpage Dialog". The address bar at the bottom shows "http://nlbawqavubis/sm100dev/ClientBody.csp?ClientId=1.20003" and "Local intranet".

Bij het Updateprofiel moet het catalogiseerprofiel opgegeven worden dat is aangemaakt in 651 / Targets

### Let op

Er moet tenminste één criterium worden opgegeven. In het voorbeeld is dat "als ISBN of ISSN is gedefinieerd".

### Verwerking

AFO 141 – maak een bewaarlijst van records in de huidige lokale database. Dit zijn de basisrecords die overschreven kunnen worden door inkomende gegevens. In plaats daarvan kan het selectieprofiel gebruikt worden om zelf een bewaarlijst aan te maken. In het voorbeeld hebben we een bewaarlijst uit database 1 en voeren we de selectie voor harvesting uit waarbij records worden opgeslagen in database 6.



Applicatie: Bibliografische beschrijvingen  
 Wachtwoord van: Tineke  
 Aantal records: 3  
 Geen filter gedefinieerd  
 Type weergave: Normaal

Gedefinieerde records van bewaarlijst 'harvest'

Nr.	Record Id	Sorteersleutel 1	Sorteersleutel 2	Sorteersleutel 3	Primaire verwoording record	Extra info
<input type="checkbox"/>	1 6.1030				A halloween mask	
<input type="checkbox"/>	2 6.1031				Hazel's happy Halloween	
<input type="checkbox"/>	3 6.1032				Tom's Halloween tricks	

De selectie wordt uitgevoerd vanaf de bewaarlijst. Het eerdere gedefinieerde selectieprofiel voor harvesting of wijzigen van records wordt gebruikt.

V-smart [ 2.0.01 ] (LogVisit - bezig) -- Webpage Dialog

### Selecteer een selectie profiel

Selectie profiel:

Gekoppeld selectie profiel:

Antwoorden opslaan als default:

<http://nlbawqavubis/sm200qa/ClientBody.csp?ClientId=1.23001> Local intranet

V-smart [ 2.0.01 ] (LogVisit - bezig) -- Webpage Dialog

### Uitvoeren selectie voor profiel 'HarvestforMarc21ISBN'

Proces commentaar

Start profiel op basis bestand

Start profiel op bewaarlijst HARVEST - 3 record(s)

Start profiel op gekoppelde bewaarlijst

Van record - tot record

Verwerk bewaarlijst in omgekeerde volgorde

Maximaal aantal treffers 9999999

Van datum 24/11/2009

Tot datum 24/11/2009

Resultaat bestand HarvestResults

Resultaat in unieke bewaarlijst bij meerdere gevallen

Gebruik datum/tijd indicatie voor het maken van de unieke bewaarlijst

Indien niet uniek, verwijder de bestaande inhoud van de bewaarlijst

Publiceer het resultaat naar file met

Resultaat commentaar

Antwoorden opslaan als default

OK  
Annuleer  
Help

http://nlbawqavubis/sm200qa/ClientBody.csp?ClientId=1.23001 Local intranet

Het resultaat is dat de basisrecords nu zijn aangevuld dan wel als nieuw record ingevoerd (afhankelijk van het profiel). In het voorbeeld gebruikten we basisrecords uit database 1 en sloegen de nieuwe records op in database 6.

Document Acties Bewerken

```

***      00000nam##2200000#a#4500
001      UkOxUb10227707
003      UkOxU
005      19960510070824.0
008      091124s   enk      000 | eng d
015 ## $a GB-8918004
020 ## $a 0859532968
082 ## $2 19
          $a 823/.914
092 ## $a D0400512031
100 ## $= Auth:495:874
          $a Meyrick, Kathryn.
245 10 $a Hazel's happy Halloween
          $c illustrated by Kathryn Meyrick
260 ## $a [Swindon]
          $b Child's Play
          $c c1987
300 ## $a [32]p.
          $b ill.
          $c 28 cm
  
```

Record editor - document '6.300228' - template 'Default-MARC21/B'

Record Edit

```

***      00727nam##2200205#a#45M0
001      UkOxUb10466009
003      UkOxU
005      19960512210621.0
008      910701s1988   enk      000 1 eng d
020 ## $a 0140508791
092 ## $a D0400902257
100 1# $= Auth:495:1053
          $a Mueller, Virginia.
245 12 $a A Halloween mask for monster /
          $c Virginia Mueller ; pictures by Lynn Munsinger
260 ## $a London :
          $b Penguin,
          $c 1988
300 ## $a [19] p. :
          $b ill. ;
          $c 18 x 23 cm
500 ## $a First published in the United States of America: Albert Whitman, 1986
500 ## $a Picture Puffins
700 1# $= Auth:495:1054
          $a Munsinger, Lynn.
  
```

Field : Leader

## Voorbeeldrapport

De harvester rapporteert wat is er is verrijkt, wat niet is verrijkt enz. Er kunnen ook bewaarlijsten geproduceerd worden van wel/niet verrijkte records. Deze bewaarlijst zijn terug te vinden in AFO 141. Het doel hiervan is: de gebruiker kan de records nakijken (zien wat er gebeurd is, omdat je dat niet weet aangezien alles volledig automatisch verloopt). De bewaarlijst van niet verrijkte records kan gebruikt worden om later handmatig bij te werken of voor een andere « harvesting sessie » (uit een andere database, of dezelfde database maar 2 maanden later).

Het rapport toont aantallen voor de diverse mogelijkheden (geladen, niet gevonden, enz.) en daarbij database en record nummer voor problemen (bijv. 6.245):

Z39.50 harvester

-----

Records processed : 100

Loaded records : 76

Not found : 22

-----

6.245

6.276

6.315

6.324

6.329

6.338

6.341

6.355

6.361

6.391

6.393

6.396

6.399

6.402

6.404

6.406

6.409

6.416

6.417

6.250

6.360

6.400

-----

More than one found : 1

6.366

-----

Errors : 1

---

6.199 : No key in local record

**Of**

Z39.50 harvester

---

Records processed : 4

Loaded records : 4

Not found : 0

More than one found : 0

Errors : 0

- **Document control - Change History**

<b>Version</b>	<b>Date</b>	<b>Change description</b>	<b>Author</b>
1.0	November 2009	creation part of 2.0 updates	