

DH – Data harvester

1 Introduction

Vubis supports an automatic harvester for bibliographic records, using Z39.50 as the harvesting protocol.

The 'harvester' automates the copy cataloguing workflow which allows you to use the Z39.50 cataloguing profiles to search Z39.50 targets and allows the merge of data into your original cataloguing records.

Below is a description of the steps you need to take before you can use this functionality.

2 Setup

A matching profile needs to be set up in AFO 114 to determine the field(s) to use for determining if records are the same.

A merging profile needs to be set up in AFO 115 to specify what must happen to fields when you merge records that are deemed to be 'the same' based on the matching profile.

Optionally create a conversion profile in AFO 134, in case the format of incoming records is different from the format used in your target database.

A standard import profile is defined in AFO 133 which specifies the Z39.50 input record type, any conversions if required, match and merge criteria, destination database and template, etc. This is linked to the previously defined matching and merging profiles (and optionally to the conversion profile).

This import profile is then specified on the definition for the Z39.50 Database in AFO 651. A Z39.50 target is created for the copy partners, and a cataloguing profile is created to link the original record key to the copy source. See AFO 651 - Z39.50 Parameters - Target Profiles - Cataloguing Profiles for details.

A selection profile is created in AFO 141 to retrieve a list of bibliographic records. There are two fields on the input form for creating a selection profile that are valid only for the bibliographic application: 'Update type' and 'Update profile'.

See the online help of the aforementioned AFO's for more detailed information on each step.

3 Workflow


The selection profile can be executed either against a predefined savelist or may be used to create a new savelist. The system will loop through the resulting savelist and process each bibliographic record as follows:

1. A Z39.50 search will be started using the distinctive search criteria within the cataloguing profile selected. This could be EAN, ISSN, ISBN or similar distinctive search criterion (key).
2. The request will generate one of three possible options:
 - no record is returned – a message is written to the report file
 - one record is returned – the bibliographic record is updated using the standard load and update tools
 - more than one record is returned – a message is written to the report file

The error / processing reports can be viewed using the standard report capabilities in AFO 642 in a standard txt format.

Example:

We have a number of skeleton records which are collected and placed into a savelist



Record	Actions	Edit
*** 008 091124s enk 000 eng d 020 ## \$a 0140508791 245 10 \$a A halloween mask \$b Virginia Mueller		

Record Actions Edit

```

***      00000nam##2200000#a#4500
008      091124s      enk      000 | eng d
020 ## $a 0859532968
245 10 $a Hazel's happy Halloween
  
```

Application Bibliographic descriptions
 User Tineke
 Number of records 3
 No filter defined
 Type of view Normal

Defined records of savelist 'harvest'

No.	Record Id	Sort key 1	Sort key 2	Sort key 3	Primary record wording	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	

We would like to let the system find the full records for us from a selected source Z39.50 target.

1. Set up a match profile in AFO 114 to specify how to handle updates to the record. In this case we will use the Marc21 020 \$a (ISBN) and 022\$a (ISSN) index entries as our match criteria:

Marc21ISBN [Bib]

Matching profile rules

No.	Index or matching file	Action 0 matches	Action 1 match	Action more matches
<input type="checkbox"/>	1 Index - ISBN(M21) [Isbn]	New	Update	To savelist
<input type="checkbox"/>	2 Index - ISSN [Issn]	New	Update	To savelist

2. Create a MERGE profile ruleset in AFO 115 to specify how to perform the record merge. In this case we are going to overwrite the existing record with the incoming record from our source:

The details are:

3. In AFO 134 create a conversion profile if required, for the mapping of data between the source or incoming record to the destination or record to be filed in the local database.

4. Create an Import profile in AFO 133. This will define the character set, format of the incoming record(s), conversion profile to be used, savelists to be used during the load process and the links for the match profile, merge profile, destination database information, conversion profile etc. Since the Harvesting protocol currently in use is Z39.50, ensure that the import format has been defined to use the proper ISO2709 format.

Input/Output profile

General **Format** **Input** Save lists Items FTP

Matching profile: Marc21ISBN

Merging profile: M21Overwrite

Rebuild relations when merging:

Database: Bibliographic (Marc21 format)

Default template: Default-MARC21/B

Maximum errors in record: 10

Maximum program errors for import: 10

Maximum program errors for load: 10

Set record modification status to: Modifications allowed

Set record delete status to: Deletion allowed

If data match rejected form: Link to main heading

Save settings:

OK
Cancel
Help

- In AFO 651 - Z39.50 create a database and database group using the Z39.50 server to be accessed for source or incoming records. On the Data Source definition, specify the Import profile created for the Harvesting (match and merge) as defined 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

Z39.50 DatabaseName: ADVANCE

Comments:

In Use?

Ask for login and password?

Group (for login):

Login name:

Password:

Repeat Password:

Supports browse (Z39.50 Scan)?

Supports 'Named result sets'?

Preferred Record Syntax: MARC21

Import profile (for conversion): HarvestM21WithOverwrite (Bib)

OK
Cancel
Help

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

- In AFO 651 - Z39.50 - Target Profile create a target or target group to be used for query.

V-smart [2.0.02] -- Webpage Dialog

Modify a Target Profile

Name:

Wording:

Comments:

Database Groups:

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

Matching Profile:

Database(s) for Matching Profile:

Auteurs - Collectivités et Congrès(Auth/278)	Bibliographic database(Bib/6)
Auteurs - Noms de personnes (Auth/277)	
Authors(Auth/4)	
Background SMART format	
DR21(Bih/21)	

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 for charact:

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

7. In AFO 651 - Targets - Cataloguing Profile create or update a cataloguing profile

V-smart [2.0.02] -- Webpage Dialog

Name:

Wording:

Comments:

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

Select your target profile group for the Z39.50 search group. Access point is the Use Attribute which will be sent to the target Z39.50 search group.

Search key defines the field(s)/Subfield(s) to be used from the source record to identify the data to be sent as a search key to the target host.

Multiple source fields/subfields can be specified as comma delimited. (i.e. enter 020/\$a,022/\$a to specify a Marc21 ISBN or ISSN field).

8. In AFO 141 - Selection create a new selection profile

The screenshot shows a web-based dialog box titled "Properties of selection profile 'HarvestForMarc21ISBN' of application 'Bib'". The dialog contains the following fields and controls:

- Profile name:** HarvestForMarc21ISBN
- Comments:** Selection profile for 'Harvester'
- File to select from:** Custom MARC 21 Bibliographic Format
- Update type:** Update bibliographic records
- Update profile:** Marc21ISBNHarvest
- Save settings:**
- Buttons:** OK, Cancel, Help

The status bar at the bottom shows the URL: <http://nlbawqavubis/sm100dev/ClientBody.csp?ClientId=1.20003> and "Local intranet".

Specify the update profile as the Cataloguing profile you have created in 651 / Targets

Note

You must have at least one criterion defined. In our case we will use "if ISBN or ISSN is defined".

To Process

AFO 141- Create a savelist of the records you have in the current local database. These are the skeleton records which can be overwritten with the incoming data. Alternately, the selection profile could be used to create its own result list. In this example, we have already created a savelist using database 1 and will be executing the selection for the harvest filing records to database 6.

Application: Bibliographic descriptions
 User: Tineke
 Number of records: 3
 No filter defined
 Type of view: Normal

Defined records of savelist 'harvest'

No.	Record Id	Sort key 1	Sort key 2	Sort key 3	Primary record wording	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	

From the savelist, execute the selection. Use the selection profile which has been previously defined for the harvesting or updating of records.

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

Select a select profile

Selection profile: HarvestforMarc21SBN

Associated selection profile:

Save settings

OK
 Cancel
 Help

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

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

Execute selection for profile 'HarvestforMarc21ISBN'

Process comments

Run profile on basic file

Run profile on savelist HARVEST - 3 record(s)

Run profile on associated savelist

From record - to record

Process savelist in reverse order

Maximum number of hits 9999999

From date 24/11/2009

To date 24/11/2009

Result file HarvestResults

Create unique file name if process is run on multiple instances

Use date/time stamp to create the unique file

If not unique, remove existing content of savelist

Publish the result to file with

Result comments

Save settings

OK
Cancel
Help

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

The result is that the skeleton record is either updated or created as a new record (depending upon the profile setup). In this example we used skeleton records from database 1 and filed new records into database 6.

Record	Actions	Edit
***		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 <i>Meyrick, Kathryn.</i>
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

Sample report

The harvester gives a report of what could be enriched, what was not enriched, etc. It also can produce savelists of the enriched/non enriched bib records. These savelists are created in AFO 141. The purpose of this is: the user can check the records (actually see what has been done, otherwise he might never know because it is fully automatic). For the bibliographic records that were not enriched, the savelist can be used later to review manually, or for another « harvesting session » (to another database, or to the same database but 2 months later).

The report shows numbers for various possibilities (loaded, not found etc.) and lists database and record number for problems (e.g. 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

Or

Z39.50 harvester

Records processed : 4

Loaded records : 4

Not found : 0

More than one found : 0

Errors : 0

- **Document control - Change History**

Version	Date	Change description	Author
1.0	November 2009	creation part of 2.0 updates	