

# DH – Moissonneuse de données

## 1 Introduction

Vubis prend en charge une moissonneuse automatique de notices bibliographiques, en utilisant Z39.50 comme protocole de moissonnage.

La « Moissonneuse » automatise la copie de la procédure de catalogage qui vous permet d'utiliser les profils de catalogage Z39.50 pour rechercher les cibles Z39.50 et permettre la fusion des données dans vos notices originales de catalogage.

Ci-dessous une description des étapes que vous devez exécuter avant de pouvoir utiliser cette fonctionnalité.

## 2 Paramétrage

Un profil de correspondance doit être paramétré dans l'AFO 114 pour désigner la (les) zone(s) à utiliser pour déterminer si les notices sont les mêmes.

Un profil de fusion doit être paramétré dans l'AFO 115 pour spécifier ce qui doit arriver aux zones lorsque vous fusionnez les notices jugées comme étant « les mêmes » d'après le profil de correspondance.

Optionnellement créez un profil de conversion dans l'AFO 134, dans le cas où le format des notices importées serait différent du format utilisé dans votre base de données cible.

Un profil d'importation standard est défini dans l'AFO 133 qui spécifie le type de notice Z39.50 entrée, toutes les conversions à effectuer, les critères de dédoublement et de fusion, la base de données de destination et la grille à utiliser, etc. Ceci est lié au dédoublement précédemment définie et aux profils de fusion (et optionnellement au profil de conversion).

Le profil d'import est alors spécifié dans la définition pour la base de données Z39.50 dans l'AFO 651. Une cible Z39.50 est créée pour les partenaires, et un profil de catalogage est défini, qui sert à lier la notice originale à la notice dérivée. Voir l'AFO 651 / Z39.50 / Profils cibles / Profils de catalogage. Voir l'AFO 651 – Paramètres Z39.50 – Profils cibles – Profils de catalogage, pour plus de détails.

Un profil de sélection est créé dans l'AFO 141 pour récupérer une liste de notices bibliographiques. Sur l'écran de saisie il existe deux zones pour créer un profil de sélection, valides seulement pour l'application bibliographique : « Type de modification » et « Modifier profil ».

Voir l'aide en ligne de l'AFO susmentionnée pour une information plus détaillée sur chaque étape.

### 3 Procédure

Le profil de sélection peut être exécuté soit sur un panier prédéfini soit utilisé pour créer un nouveau panier. Le système procédera par itérations dans le panier de résultat et traitera chaque notice bibliographique de la manière suivante :

1. Une recherche Z39.50 est effectuée, qui utilise les critères de recherche (EAN, ISSN, ISBN, etc.) définis dans le profil de catalogage sélectionné.. Cela peut être EAN, ISSN, ISBN ou un critère de recherche distinctif similaire (clé).
2. L'un des trois cas de figure suivants se produira :
  - aucune notice n'est retournée – l'événement est noté dans le compte rendu
  - une notice est retournée – la notice bibliographique est mise à jour à l'aide des outils de chargement et de mise à jour standard
  - plusieurs notices sont retournées – l'événement est noté dans le compte-rendu

Les rapports d'erreur / de traitement peuvent être visualisés en utilisant les possibilités de rapport dans l'AFO 642 dans un format txt standard.

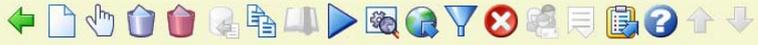
#### Exemple :

Nous avons un certain nombre de notices minimum collectées et placées dans un panier



Notice	Actions	Editer
***	p0000nam##2200000#a#4500	
008	091124s enk 000   eng d	
020 ## \$a	0140508791	
245 10 \$a	A halloween mask	
	\$b Virginia Mueller	

Notice	Actions	Editer
***	00000nam##2200000#a#4500	
008	101214s      enk      000   eng d	
020 ## \$a	0859532968	
245 10 \$a	Hazel's happy Halloween	

↩

Application	Notices bibliographiques
Mot de passe de	Tineke
Nbre d'enregistrements	3
Aucun filtre n'a été défini	
Type d'affichage	Normal

Enreg. définis pour le panier 'harvest'

N°	Id enreg.	Critère de tri 1	Critère de tri 2	Critère de tri 3	Liste des documents trouvés	Info supplémentaire
<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	

Nous souhaiterions laisser le système trouver pour nous les notices complètes à partir d'une cible source Z39.50 sélectionnée.

1. Paramétrer un profil de correspondance dans l'AFO 114 pour spécifier comment traiter les mises à jour pour la notice. Dans ce cas nous utiliserons les entrées des index MARC 21 020 \$a (ISBN) et 022\$a (ISSN) comme critère de profil de dédoublement :

↩

Marc21ISBN [Bib]

Règles du profil de dédoublement

No.	Index ou fichier de dédoublement	Action rien trouvé	Action 1 notice trouvée	Action plusieurs notices trouvées
<input type="checkbox"/>	1 Index - ISBN(M21) [Isbn]	Nouvelle notice	Mise à jour	Garder dans liste
<input type="checkbox"/>	2 Index - ISSN [Issn]	Nouvelle notice	Mise à jour	Garder dans liste

2. Créer l'ensemble des règles d'un profil de FUSION dans l'AFO 115 pour préciser comment effectuer la fusion de notices. Dans ce cas nous allons écraser la notice existante avec la nouvelle notice issue de notre source :

Les détails sont :

3. Dans l'AFO 134 créer un profil de conversion si nécessaire, pour la conversion des données depuis la source ou la notice entrante vers la destination ou la notice à enregistrer dans la base de données local.
  
4. Créer un Profil d'import dans l'AFO 133. On y définira le jeu de caractères, le format de la (des) nouvelle(s) notice(s), le profil de conversion à utiliser, les paniers à utiliser pendant le traitement de chargement et les liens pour le profil de correspondance, le profil de fusion, les détails de la base de données de destination, le profil de conversion, etc. Puisque le protocole de moissonnage habituellement utilisé est Z39.50, assurez-vous que le format d'import a été défini pour utiliser le format approprié ISO2709.

**Profil import/export**

Général Format Import Paniers Exemples FTP

Profil de recherche de doublons: Marc21ISBN

Profil de fusion: M21Overwrite

Reconstruire relations lors de la fusion:

Base de données: Bibliographic database

Grille de saisie par défaut: Default-MARC21/B

Maximum erreurs pour une notice: 10

Maximum erreurs de logiciel pour l'importation: 10

Maximum erreurs de logiciel pour chargement: 10

Statut modification notice devient: Modifications autorisées

Statut suppression notice devient: Suppression autorisée

Si données égales à forme rejetée: Lier à forme retenue

Stocker réponses comme val. par défaut:

OK  
Annuler  
Aide

5. Dans l'AFO 651 – Z39.50 créer une base de données et un groupe de bases de données en utilisant le serveur Z39.50 auquel la source ou les nouvelles notices doivent accéder. Sur la définition Donnée Source, spécifiez le profil d'import créé pour le Moissonnage (correspondance et fusion) comme défini dans l'AFO 133.

V-smart [ 2.0 ] -- Webpage Dialog

Nom: ForHarvest

Formulation pour l'utilisateur: UseForHarvesting

Adresse Internet: 163.1.13.199

No de Port: 210

Nom de la base de données (Z39.50): System

Commentaire:

Utilisé?

Demander nom et mot de passe:

Id groupe (pour login):

Nom Login:

Mot de passe:

Repetier mot de passe:

Peut feuilleter (Scan)?

Noms pour ensembles de résultats:

Syntaxe préférée: MARC21

Profil de recherche (pour l'importation): harvestM21overwrite (Bib)

Profil d'exportation (pour le catalogage collectif):

Les vedettes de feuilletage ont le jeu de caractères décrit dans le:

OK  
Annuler  
Aide

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

- Dans l'AFO 651 – Z39.50 – Profils cibles créez une cible ou un groupe cible à utiliser pour l'interrogation.

**Modifier un profil de cible**

Nom: GroupDatabase

Formulation: Group of Databases

Commentaire:

Groupes de bases de données: [List of databases including Ivubis-Local Marc21, ANTWERP University, etc.]

Profil de dédoublement: [Dropdown menu]

Base(s) de données pour le profil de correspondance: [List of databases including Background database 1 for VSMANT1302(Bib/21), etc.]

Utiliser la base de données pour les caractéristiques: Bibliographic database(Bib/6)

Première recherche : trop de résultats quand les résultats >= 10000

Première recherche : maximum à trouver >= 20

Première recherche : par présentation >= 10

Time out après: [Field]

Buttons: OK, Annuler, Aide

URL: http://nlbawqavubis/sm200qa/ClientBody.csp?ClientId=1,23001

- Dans l'AFO 651 – Cibles – profil de catalogage créez ou mettez à jour un profil de catalogage

**Modifier un profil de cible**

Nom: Marc21ISBNHarvest

Formulation: Marc21ISBNHarvest

Commentaire:

Profil de cible: GroupDatabase

Index: ISBN

Définir la clé de recherche en utilisant une ou plusieurs combinaisons de zones/sous-zones: 020\$a,022\$a

Strip punctuation?

Buttons: OK, Annuler, Aide

URL: http://nlbawqavubis/sm200qa/ClientBody.csp?ClientId=1,23001

Sélectionnez votre profil groupe cible pour le groupe de recherche Z39.50. Le point d'accès est l'Attribut d'utilisation qui sera envoyé vers le groupe de recherche cible Z39.50.

La clé de recherche définit la (les) zone(s)/Sous-zone(s) à utiliser dans l'enregistrement source pour identifier les données à envoyer en tant que clé de recherche vers l'hôte cible.

Plusieurs zones/sous-zones source peuvent être spécifiées, séparées par une virgule (par exemple saisir 020\$a,022\$a pour spécifier une zone ISBN ou ISSN Marc21).

#### 8. Dans l'AFO 141 – Sélection, créez un nouveau profil de sélection

Propriétés du profil de sélection 'HarvestforMarc21ISBN' ou application 'Bib'

Nom du profil: HarvestforMarc21ISBN

Commentaire: Selection profile for harvesting

Sélections à partir du fichier: Format as defined by the MARC21 comitte

Type de modification: Mise à jour des notices bibliographiques

Modifier profil: iMarc21ISBNHArvest

Stocker réponses comme val. par défaut:

OK, Annuler, Aide

Spécifiez comme profil à modifier le Profil de catalogage que vous avez créé dans l'AFO 651 / Cibles

#### Remarque

Vous devez avoir au moins un critère défini. Dans notre cas nous utiliserons « si ISBN ou ISSN est défini ».

#### Exécution

AFO 141 – Créez un panier des notices que vous avez dans la base de données locale courante. Celles-ci sont les notices minimum qui peuvent être écrasées par les nouvelles données. Sinon, le profil de sélection peut être utilisé pour créer sa propre liste de résultats. Dans cet exemple, nous avons déjà créé un panier en utilisant la base de données 1 et exécuterons la sélection pour le moissonnage des notices enregistrées vers la base de données 6.

Application: Notices bibliographiques  
 Mot de passe de: Tineke  
 Nbre d'enregistrements: 3  
 Aucun filtre n'a été défini  
 Type d'affichage: Normal

Enreg. définis pour le panier 'harvest'

N°	Id enreg.	Critère de tri 1	Critère de tri 2	Critère de tri 3	Liste des documents trouvés	Info supplémentaire
<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	

A partir du panier, exécutez la sélection. Utilisez le profil de sélection qui a été précédemment défini pour le moissonnage ou la mise à jour des notices.

Choisir un profil de sélection

Profil de sélection: HarvestforMarc21ISBN

Profil de sélection associé: [vide]

Stocker réponses comme val. par défaut:

OK  
Annuler  
Aide

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

V-smart [ 2.0 ] -- Webpage Dialog

### Exécuter la sélection pour le profil 'HarvestforMarc21ISBN'

Commentaire activé

Exécuter le profil à partir de la base de données

Exécuter profil à partir d'un panier HARVEST - 3 notice(s)

Exécuter le profil sur le panier associé

De l'enreg. - à l'enreg.

Traitement de panier dans l'ordre inverse

Nombre maximal d'occurrences 9999999

A partir de date 14/12/2010

Jusqu'à date 14/12/2010

Fichier de résultats

Résultat dans panier unique lors de cas multiples

Ajouter date/heure pour créer un fichier unique

Si non unique, supprimer le contenu du panier

Editer le résultat avec

Commentaire résultat

Stocker réponses comme val. par défaut

OK  
Annuler  
Aide

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

Le résultat est que la notice minimum est soit mise à jour soit créée comme nouvelle notice (selon le paramétrage du profil). Dans cet exemple nous avons utilisé les notices minimum de la base de données 1 et enregistré les nouvelles notices dans la base de données 6.

Notice	Actions	Editer
***		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

## Sample report Extrait de rapport

Le moissonneur fournit un rapport de ce qui pourrait être enrichi, de ce qui n'a pas été enrichi, etc. Il peut également produire des paniers des notices bibliographiques enrichies/non enrichies. Ces paniers sont créés dans l'AFO 141. Le but de l'opération est de permettre à l'utilisateur de vérifier les notices (en réalité voir ce qui a été fait, sinon il ne le saurait jamais car l'opération est entièrement automatique). Pour les notices bibliographiques qui n'ont pas été enrichies, le panier peut être utilisé plus tard pour une révision manuelle, ou pour une autre « session de moissonnage » (vers une autre base de données, ou vers la même base de données mais 2 mois plus tard).

Le rapport affiche les chiffres pour diverses éventualités (chargé, non trouvé, etc.) et liste les base de données et les numéros de notice en cas de problèmes (ex. : 6.245) :

moissonneur Z39.50

-----

Notices traitées: 100

Notices chargées: 76

Non trouvées: 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

---

Plus d'une trouvée : 1

6.366

---

Erreurs : 1

---

6.199 : Aucune clé dans une notice locale

**Ou**

Z39.50 harvester moissonneur Z39.50

---

Notices traitées : 4

Notices chargées : 4

Non trouvées : 0

Plus d'une trouvée : 0

Erreurs : 0

- **Document control - Change History**

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