AFO 133 - Input /output profiles

133.1 Introduction

Use this AFO to create and maintain profiles for importing records into Cataloguing databases (via AFO 131), and exporting records from Cataloguing databases (via AFO 132).

133.1.1 File formats

The import/export module can handle multiple digital file formats. Below are descriptions of several varieties of raw data.

1. "Delimited"

A delimited file contains data separated by specific characters. Usually more than one separator is used, at least one for separating data fields and one for separating records.

	A	B	C	D	E	F
1	ID	Code	Author	Title	Additional data	Publication info
2	2	aa02	Mickley, H.S. / Sherwood, T.K. / Reed,	Applied mathematics	-	New York, Mc-Graw-Hill,
3	3	aa03	University of Delaware	The application of dif	-	Newark, 1957
4	4	aa04	Wylie, C.R.	Advanced engineerin	eerste editie 1951	New York, McGraw-Hill,
5	5	aa05	Rodguin, N.M. / Rodguina, E.N.	Consecutive chemica	Oorspronkelijk in H	New York, Nostrand, 196
6	6	aa06	Dermijn, J.F./Bock, G.A. de/Gelder, T.	Wiskunde in de proce	Uitgave ten dienst	Culemborg, Stam, 1968
7	8	aa08	Blij, F.van der/ Tiel, J. van	Infinitesimaalrekening	-	Utrecht, Spectrum, 1969
8	9	aa09	Linden, C. van der	Wiskundig vademecu	-	Utrecht, Spectrum, 1969
9	10	aa10	Lehmann, H.	De rekenliniaal.	Oorspronkelijke tit	Utrecht, Spectrum, 1969
10	11	aa11	Graham Flegg, H.	Schakelalgebra.	Oorspronkelijke tit	Utrecht, Sprectrum, 1969
11	13	aa14	Ahten, H./Bruhn J.	Elektronisch rekenen	Oorspronkelijke tit	Deventer, Kluwer, 1975.
12	15	aa16	Dence, J.B.	Mathematical technic	-	New York, Wiley, 1975.
13	17	aa18	Russel, T.W.F./Denn, M.M.	Introduction to chemic	-	New York, Wiley, 1972.
14	18	aa19	Teller, O	Vademecum van de	achtste druk.	Utrecht, Het Spectrum. 19
15	19	ad01	Imperial Chemical Industries	Statistical methods in	4th revised edition	London, Longman, 1972
16	21	ad03	Wijvekate, M.L.	Verklarende statistiel	-	Utrecht, Het spectrum, 19
17	22	ad04	Hemelrijk, J.	Statistiek, te pas en te	-	Deventer, Kluwer, 1972
18	23	ad05	Sachs, L.	Statistische Methode	zweite neubearbei	Berlin, Springer, 1972.
19	24	ad06	Grinten, P.M.E.M. van der/Lenoir J.M.N	Statistische procesbe	-	Utrecht, Spectrum, 1973.
20	25	ad07	American Society for Testing and Mate	Manual on statistical	Manul on statistica	Philadelphia, 1975
21	26	ad08	International Organization for Standar	Statistical methods; h	Texts for ISO stand	Geneva, 1979
22	27	ad10	Spiegel, M.R.	Theory and problems	-	New York, McGraw-Hill,
23	28	ad11	Boon, K.L.	Practische Statistiek	-	Leiden, Spruyt, Mantege
24	29	ad12	Hooke, R.	How to tell the liars fro	-	New York, Marcel Dekke

In the example we see an Excel sheet with bibliographic data. From Excel (or Access) it is relatively simple to save the delimited data.

Example:

A file with delimited data (with ";" as delimiter):

Title; author; place of publication; publisher; etc.

The import module processes the data by column as they were saved in Excel or Access, where column 4 in the example is the title, column 3 the author etc.

The only restriction the import module has is that each column is a filed. You cannot spread data from one column over multiple fields. In practice you often see all authors of a title in one column. For a successful import the primary author and all subsequent authors should be put in separate columns.

2. "Tagged"

A tagged format implies each data field is preceded by a label that allows the import module to determine what type of data is in it.

An example of a pure tagged format is:

```
IT 94 D 0001|Diaserie|OWP 94 D 0001, thema: kunst, kast 41
$
DN ABC97LM06577
TY LM
LA Dutch
MT Slides
SD BASISONDERWIJS BOVENBOUW
SD BASISONDERWIJS MIDDENBOUW
TI HET GALANTE LEVEN
CT 9 p. : + tekstboek
SI OWP 94 D 0110, thema: kunst, kast 41
DE SCHILDERKUNST 18DE EEUW
DE PORTRET
DE KUNSTBESCHOUWING
IT 94 D 0110|Diaserie|OWP 94 D 0110, thema: kunst, kast 41
$
DN ABC97AV06773
TY AV
LA Dutch
MT Slides|
SD BASISONDERWIJS BOVENBOUW
TI NATUURLIJKE GRONDSTOFFEN 2 (o.a. metalen-zout)
CT 20 p. : + 1 tekstboek
SI OWP 83 D 0600, thema: industrie, kast 41
DE GRONDSTOFFEN
DE IJZERERTS
DE METALEN
DE ERTSEN
DE GOUD
DE ZILVER
DE ZOUT
IT 83 D 0600|Diaserie|OWP 83 D 0600, thema: industrie, kast 41
```

The first two letters on each line determine the type of field. So: LA = language, MT = material type, TI = title, DE = descriptor, IT = item details, etc. After the space you have the contents of the filed. The dollar sign means 'end of record'. This must be an ASCII text file which will look like the above example when opened with Notepad.

A tagged file can also contain text of variable length. Like:

```
shelfmark|CD114
category code|C
category description|CD
division code|ALG
division description|Algemeen
sequence number|011466
author|
title|Commercials Centraal Beheer Achmea [DVD]
series title|
publisher|Centraal Beheer Achmea
place of publication|Apeldoorn
year of publication|2004
extra division1|C
descriptor1|AUDIOVISUELE MEDIA
descriptor2|RECLAME
descriptor3|PLAATS DELICT
descriptor4|
descriptor5|
descriptor6|
descriptor7|
descriptor8|
descriptor9|
descriptor10|
descriptor11|
descriptor12|
annotation|distr. Centraal Beheer Achmea (36 min.)
```

In this case you need a special character to mark the end of the tag and the beginning of the data. IN the example this is a pipe, "|". A space would not be sufficient, because the data (field content) can also contain spaces.

For both delimited and tagged formats you must determine what the separator characters are going to be for fields and records. The import module needs to know the ASCII value of those characters.

3. "XML"

Exporting and importing records in XML format is incorporated in the standard facilities for data export (AFO 132) and import (AFO 131). It applies to both bibliographic documents and authority records. Both export and import use the profiles created and modified through AFO 133, which in its turn optionally use conversion profiles defined through AFO 134. All this implies that XML records are imported in exactly the same manner as ISO2709, delimited or tagged records.

If you have a database using one of the three most common formats used within the application (MARC21, UNIMARC and Smart), you have the following possibilities:

- Export and import of MARCXML records.

- Use an import/export profile that uses the file format MARCXML.

- Export and import of Dublin Core in XML format.

- Use an import/export profile that uses the file format XML Dublin Core and use as a conversion profile one of the predefined XML Dublin Core conversion profiles.

- Export and import of other XML formats.

- There are multiple possibilities here:

- Create your own conversion profile in AFO 134 and create an import/export profile in AFO 133 that uses "XML Generic" as file format. Under "record delimiter", fill in the name of the element that indicates the start of a new record.

- Import/export ISO 2709 records and use a conversion tool outside of Vubis Smart.

- Use techniques outside of Vubis Smart to convert MARCXML or DC-XML to or from the desired format. Standard techniques such as XSLT can be used to create such conversions.

An example of Dublin Core XML:

```
<?xml version="1.0" ?>
- <dc xmlns="http://purl.org/dc/elements/1.1/">
  <title>Arithmetic /</title>
  <creator>Sandburg, Carl, 1878-1967.</creator>
  <creator>Rand, Ted, ill.</creator>
  <type />
  <publisher>San Diego :Harcourt Brace Jovanovich,</publisher>
  <date>c1993.</date>
  <language>eng</language>
  <description>A poem about numbers and their characteristics. Features
anamorphic, or distorted, drawings which can be restored to normal by
viewing from a particular angle or by viewing the image's reflection in
the provided Mylar cone.</description>
  <description>One Mylar sheet included in pocket.</description>
  <subject>Arithmetic</subject>
  <subject>Children's poetry, American.</subject>
  <subject>Arithmetic</subject>
  <subject>American poetry.</subject>
  <subject>Visual perception.</subject>
  </dc>
```

133.1.2 Character set conversion

Below are some notes on the conversion and management of MARC21 records in the ALA character set when imported into a Unicode Marc record.

Incoming records in the ALA character set can be fully converted to Unicode, based on the conversion rules described by Library of Congress at
http://www.loc.gov/marc/specifications/specchartables.html

This allows fully for the Latin and Extended Latin characters, for Hebreew, Cyrillic, Arabic, Greek and East Asian character repertoires.

It is based on the MARC21 Character coding scheme (Field 09 in leader). A "format test" can be added for this field in the MARC21 tables.

GENERAL		
1 Comments and notes		
2 Data type	Text	
3 Definition	A one-character code that identifies the character coding scheme ...	
4 Examples		
Identification	***!10**	
5 In use / active	Yes	
6 Punctuation after subfield		
7 Punctuation before subfield		
8 Label	Character coding scheme	
9 Language dependant	No	
10 Mandatory	No	
11 Maximum number	999	
12 Minimum number	0	
13 Standard subfield	Yes	
14 Unique value	No	
15 Wording	Character coding scheme	
FORMAT		
16 Right justified	No	
17 Default		
18 Fill character	#	
19 Fixed length	Yes	
20 Format pattern		
21 Format test	Marc21Leader	
22 Generated data		
23 Maximum length	1	
24 Minimum length	1	
25 Modification allowed	Yes	
26 Valid values		

When this is defined, records imported to a Unicode based database will have this field in the leader set to "a" to indicate that this is a Unicode representation. This ensures that the Marc records conform to the standard.
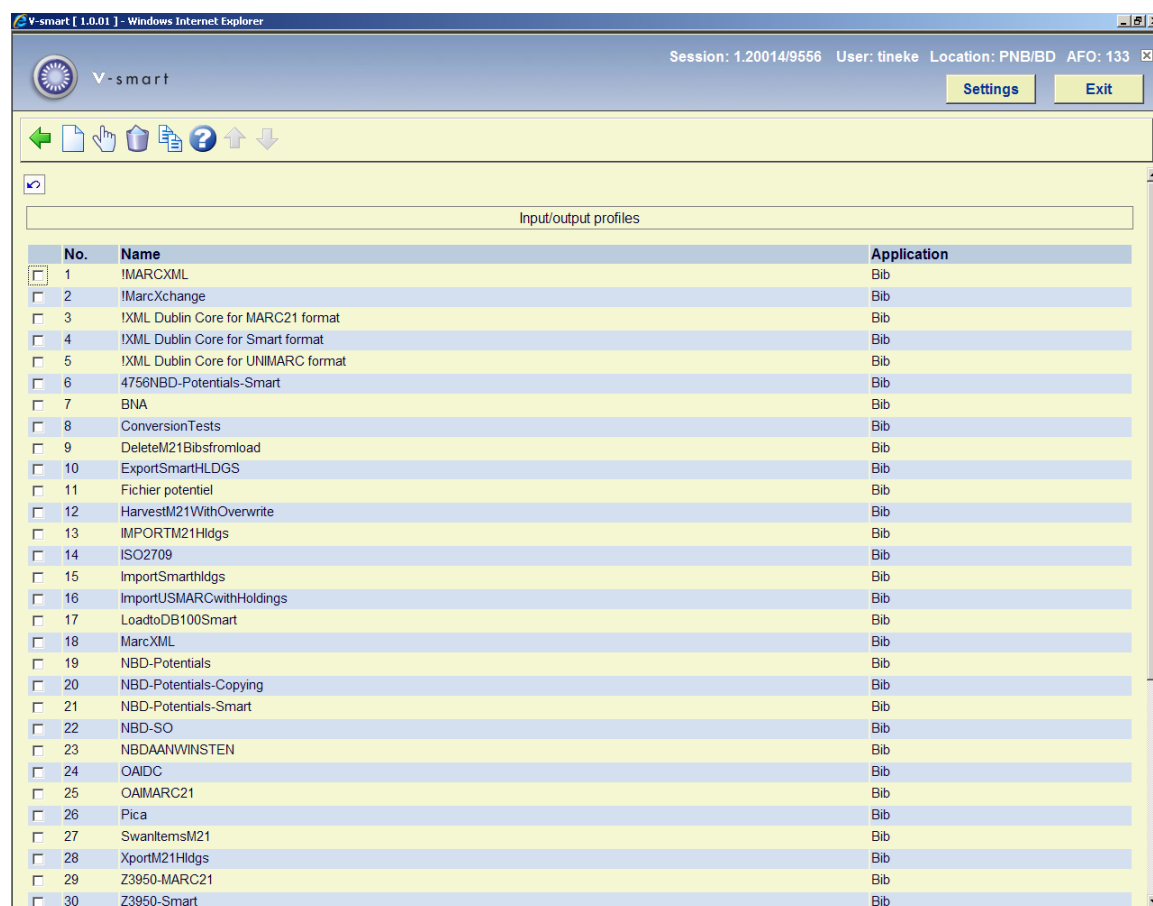
A setting in the import loader tells the system to use the incoming character set of the MARC record to determine its character set (see below under Tab General).
If this is set then the system will convert from ALA to Unicode or leave the data as Unicode as appropriate.

Note

Libraries using this option are warned to test their incoming data, as obviously the system cannot cross check the integrity of data supplied by some other supplier.

133.2 Definitions in AFO 133

After you have chosen this AFO an overview screen with existing profiles will be presented. You can edit existing profiles on this screen or create new ones.

Options on the screen

New item: click this icon to create a new profile. You will first be presented with an input form where you must enter the profile name and select an application (bibliographic or authority).
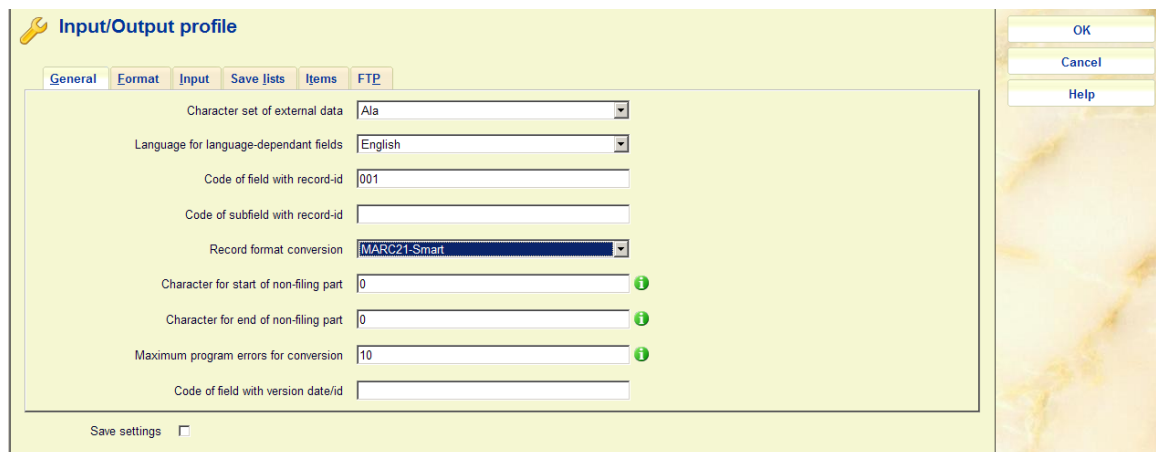
View/modify item properties: select an existing profile and click on this icon to modify the profile.

Delete item: select an existing profile and click on this icon to delete the profile.

Copy: select an existing profile and click on this icon to create a new profile by copying the existing one.

After selecting an existing profile or creating a new one, you are presented with an input form with 5 tabs. All tabs must be filled out. They are described in more detail below.

133.2.1 Tab 'General'

This tab contains the general characteristics of the input/output file.

Fields on the screen

Character set of external data: Select a standard character set from the dropdown list. If you are not sure, just choose one and look at the result in a test load. If the result is incorrect (for example strange characters appear) you can adjust by choosing another character set. For the ALA to Unicode conversion for MARC records you must select the set *From Marc21 record loader*.

Language for language-dependant fields: Select a language form the dropdown list. Usually this is set to <all languages>.

Code of field with record ID: In the example this is 1. This means the loader will assume the first field will contain a record number. In MARC21 tag 001 contains the control number. This can then be used to store separately. Because Vubis Smart has its own record ID numbers, the external record ID is not used as such. But it can be useful to retain the information in the record

Code of subfield with record ID: Here you can specify the specific subfield that contains a record ID (optional).

Record format conversion: This is one of the most critical steps. Select a profile form the dropdown list. The profiles are defined in AFO 134. You can do that later and then come back to this tab to add it.

Character for start of non-filing part: To denote where the non-filing character must be positioned in Vubis Smart. Within the Smart format you can use this to specify a certain part of the title must not be indexed. Enter the ASCII value of the non-filing character.

Character for end of non-filing part: See above.

Maximum program errors for conversion: When the process reaches the number specified here, the loading will be stopped. The system checks such details as valid fields for the format as specified in AFO 151. When for instance 'language' is a mandatory field for the format, but the incoming records contain no language then this will be seen as an error.

Code of version date/id: This allows you to specify in which field the record "version identifier" may be found. This is typically set in tag 005 for Marc21 and Unimarc.
The field specified here MUST be defined in the format described. Specifying this field causes the system to record this "version date" in the record updated in the database. There is no validation of the data entered into this field. The setting of this field is optional.

133.2.2 Tab 'Format'

This tab allows for specification of the various delimiters.

Fields on the screen

File format: Select a file format from the dropdown list.

Record delimiter: Enter the ASCII value of the separator character. Say you import an Excel file. When exporting the Excel file to a text file add the delimiters (like comma or semicolon).

Field delimiter: See above.

Subfield delimiter: See above. Subfields are not recognised in delimited files.

Tag delimiter: See above. This can be the space (ASCII value 32) between tag number and data.

Tag length: For MARC21 this should be 3 for instance.

Number of delimited fields: To define the maximum number of delimited fields. Only necessary when not all lines have the same number of delimiters. This occurs rarely.

Fixed field map: This is only used when all data can be found on fixed positions in a file.

Quotes around data: denotes whether or not the loader must take quotes into account.

133.2.3 Tab 'Input'

This tab offers the opportunity to add matching and merging profiles.

Fields on the screen

Matching profile: Select a profile from the dropdown list. These are used to specify what must happen when records are loaded that already exist in the database. You can for instance specify that the incoming record prevails. You need to create the matching and merging profiles in AFO 114 and 115.

Merging profile: See above.

Rebuild relations when merging: When records are merged, normally the relations are rebuilt. This can be very time consuming and sometimes you know beforehand this is not necessary. So you can switch this off.

Please note

The default is for this option to be switched ON. You may not switch this off without consulting the helpdesk first.

Database: When you have multiple databases, you need to specify in which database the records must be loaded. Select one from the dropdown list.

Default template: Select one from the dropdown list (the templates must be defined in AFO 153, 154).

Maximum errors: these three fields determine fault tolerance.

Record status: Choose the correct statuses from the dropdown list.

If data match rejected form: When records are loaded through AFO 131 with the record loader it may occur that the contents of an authority controlled bibliographic field match with an existing rejected form in the related authority database. Here you can specify what must happen in such a case. Three different options can be chosen if the data match a rejected form:

- Create new main heading: The rejected form is removed from the database, and instead a new main heading will be created

- Discard data: The data are removed from the import record – this is the current situation

- Link to main heading: The field is linked to the main heading of the rejected form

In all three cases a warning message will be added to the loader report.

133.2.4 Tab 'Savelists'

This tab allows various settings for creating savelists of not loaded records, loaded records, etc.

Specify whether the savelist must be created for a specific user (* means current user). You can select existing savelist names or type in new names. And you can choose to append the date to the savelist name.

133.2.5 Tab 'Items'

On this tab you can create savelists for imports that include items. By entering a period and setting the savelist to read-only you can eventually delete these items from the system. These options are only valid for import. The last option is only valid for export, here you determine from which institution(s) you export item data.
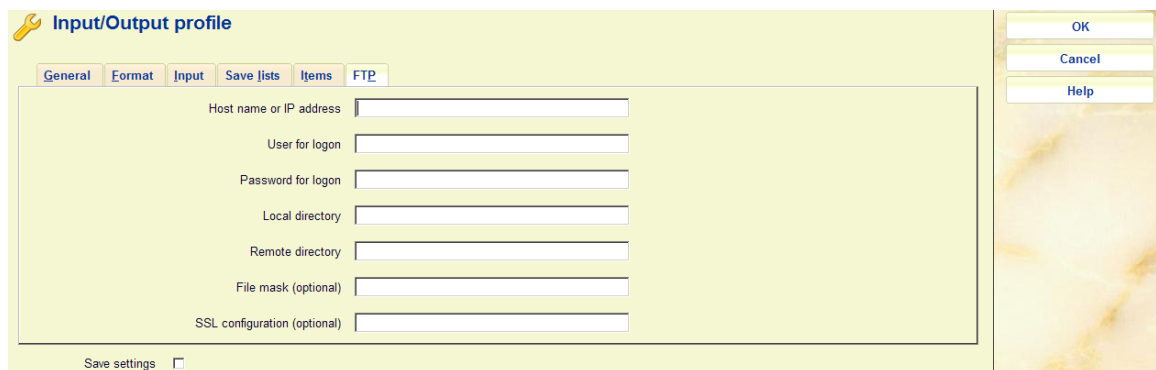
Note

Items will be exported with a status (if assigned) as defined in AFO 481 – Loan status codes – Tab General, under 'Circulation status'.

133.2.6 Tab 'FTP'

The record loader can automatically download files through FTP before the load starts. This is especially useful for automating the common workflow of

1. file creation on remote system

2. file download to local system

3. file import on local system

This tab offers the opportunity to define FTP settings. If a host name or ip-address is defined, the loader will try to connect and download files to the local system before it starts to process records.

Fields on the screen

Host name or IP address: the name or address of the remote system from which the files are downloaded.

User for logon: the ftp user name for logon on the remote system

Password for logon: the ftp user password for logon on the remote system

Local directory: the directory where the downloaded files must be saved

Remote directory: the directory that must be searched for files to be downloaded. *Note that the path is relative to the home directory of the ftp user!*

File mask (optional): optional file mask for the search in the remote directory.

SSL configuration (optional): optional SSL configuration to use for the connection.

- **Document control - Change History**

Version	Date	Change description	Author
1.0	April 2008	creation	
2.0	September 2009	various corrections; new tab for FTP settings part of 2.0 updates	
3.0	April 2010	New options for management of MARC21 Unicode vs ALA character sets part of 2.0.06 updates	