

Générateur de notices au format Unimarc pour MEDICI.TV

[Pré-requis](#)

[Spécifications fonctionnelles](#)

[Structure du "Record Label" ou "Leader"](#)

[Champs 00x : bloc de l'identification](#)

[Champ 001](#)

[Champ 003](#)

[Champ 005](#)

[Champs 1xx : bloc des informations codées](#)

[Champ 100](#)

[Champ 101](#)

[Champ 135](#)

[Champs 2xx : bloc des informations descriptives](#)

[Champ 200](#)

[Champ 210](#)

[Champ 215](#)

[Champs 6xx : indexation matière et histoire bibliographique](#)

[Champ 608](#)

[Champs 7xx : bloc des responsabilités](#)

[Champ 700](#)

[Champ 701](#)

[Champ 702](#)

[Champ 712](#)

[Champs 8xx : données internationales](#)

[Champ 856](#)

[Librairie python UNIMARC/MARC21](#)

[Lancement du programme](#)

[Programme de génération des notices](#)

[Données en entrées](#)

[Génération des notices](#)

[Date de création](#)

[Algorithme](#)

[Structure d'une notice](#)

[Structure d'un champ](#)

[Blocs d'informations codées \(100x\) - descriptives \(200x\) - Autres champs uniques](#)

[Informations casting \(700x\)](#)

[Vérification des notices](#)

[Installation](#)

[Procédure de vérification](#)

[Lancer MarcEdit](#)

[Convertir un fichier *.mrc en fichier *.mrk](#)

[Vérifier le contenu du fichier généré](#)

Pré-requis

Documentation sur le format Unimarc:

http://www.bnf.fr/fr/professionnels/anx_formats/a.unimarc_manuel_format_bibliographique.html#SHDC__Attribute_BlocArticle8BnF

Spécifications de conversion du format Unimarc ou format Marc21:

<http://www.loc.gov/marc/unimarctomarc21.html>

Spécifications fonctionnelles

Structure du “Record Label” ou “Leader”

Positions 0-4 : longueur de la notice (calculé par le programme). 5 chiffres complétés à gauche si nécessaire.

Position 5 : statut de la notice. Valeur recommandée : n pour “New record”

Position 6 : type de notice. Valeur recommandée : g pour “Projected medium”

Position 7 : niveau bibliographique. Valeur recommandée : m pour “Monograph/item”

Position 8 : non renseigné

Position 9 :

Position 10 : longueur des indicateurs. ‘2’ par défaut

Position 11 : longueur des identifiants de sous-champs (exemple: \$a). ‘2’ par défaut

Positions 12-16 : adresse de la zone de données. Calculé par le programme.

...

Position 20 : Nombre de caractères dans la longueur d’un champ. ‘4’ par défaut.

Position 21 : Longueur de la position du caractère de début d’un champ. ‘5’ par défaut.

Positions 22-23 : non renseigné

En résumé, le record label par défaut compilé pour Medici est le suivant:

LDR= ngm 22 9 450

Champs 00x : bloc de l'identification

Ce bloc contient les numéros identifiant la notice ou la ressource décrite dans cette notice.

Champ 001

Définition: identifiant de la notice.

Valeur par défaut : MEDICI-xxx ou xxx est l'identifiant de base de données de la notice.

Référence : http://multimedia.bnf.fr/unimarcb_trad/B001-6-2010.pdf

Champ 003

Définition: identifiant pérenne de la notice

Valeur par défaut : http://www.medic.tv/#?. A compléter par le chemin complet vers la notice.

Référence : http://multimedia.bnf.fr/unimarcb_trad/B003-6-2010.pdf

Champ 005

Définition: identifiant de la version

La valeur est la date de création de la notice au format

Année/Mois/Jour/Heure/Minutes/Secondes.

Référence : http://multimedia.bnf.fr/unimarcb_trad/B005-6-2010.pdf

Champs 1xx : bloc des informations codées

Ce bloc contient des données codées de longueur fixe.

Champ 100

Définition: données générales de traitement

Valeur : Date et niveau unicode.

Il s'agit d'une liste de valeurs sur 2 caractères correspondant aux positions 26-29

Référence: http://multimedia.bnf.fr/unimarcb_trad/B100-6-2010.pdf

Champ 101

Définition: Langue de la ressource

Valeur : anglais par défaut

Référence : http://multimedia.bnf.fr/unimarcb_trad/B101-6-2010.pdf

Evolutions : Medici devranormaliser les langues (supports, livrets, sous-titres....).

Champ 135

Définition : Ressources électroniques.

Valeurs : "son", "système en ligne", "couleur", "Autre dimension" et "support contenant du son"

Référence : http://multimedia.bnf.fr/unimarcb_trad/B135-6-2010.pdf

Champs 2xx : bloc des informations descriptives

Champ 200

Définition: Titre et mention de responsabilité

Valeur : titre de l'enregistrement

Référence : http://www.bnf.fr/documents/B200_6_2011.pdf

Champ 210

Définition : publication, production, diffusion

Valeur : mention de copyright

Référence : http://multimedia.bnf.fr/unimarc_b_trad/B210-6-2012.pdf

Champ 215

Définition : description matérielle

Valeur : Ressource online et format de l'image.

Référence : http://multimedia.bnf.fr/unimarc_b_trad/B215-6-2011.pdf

Champs 6xx : indexation matière et histoire bibliographique

Champ 608

Définition: Vedette matière - Forme, genre ou caractéristiques matérielles

Valeur : Archives

Référence : http://multimedia.bnf.fr/unimarc_b_trad/B608-6-2013.pdf

Champs 7xx : bloc des responsabilités

Règles à respecter :

- s'il y a plusieurs compositeurs avec le même niveau de responsabilité intellectuelle : on les met en 701. S'il y en a un qui se détache on met 700 et les autres en 702.
- Il est possible de cumuler 701 et 702.
- Il ne peut pas y avoir plusieurs 700 (auteur principal) / 710 (collectivités principales)
- Il ne peut pas y avoir 700 et 710

Champ 700

Définition: Nom de personne - responsabilité principales

Valeur : voir le fichier d'extraction.

Référence : <http://www.bnf.fr/documents/B700-6-2010.pdf>

Champ 701

Définition: Nom de personne - responsabilité secondaire.

Valeur : voir le fichier d'extraction. \$4 contient le code fonction 230 qui correspond aux compositeurs.

Références :

- http://multimedia.bnf.fr/unimarc_b_trad/B702-6-2010.pdf
- http://multimedia.bnf.fr/unimarc_b_trad/BAnnCparcodes-6-2011.pdf pour les codes fonctions

Champ 702

Définition : Nom de personne - responsabilité secondaire

Valeur : voir les compositeurs, artistes dans l'extraction. \$f contient la date de naissance de la personne.

Référence : http://multimedia.bnf.fr/unimarc_trad/BAnnCparcodes-6-2011.pdf

Champ 712

Définition : Nom de la collectivité - responsabilité secondaire

Valeur : Orchestre

Référence : http://multimedia.bnf.fr/unimarc_trad/B712-6-2011.pdf

Champs 8xx : données internationales

Champ 856

Définition : Adresse électronique et mode d'accès

Valeur : \$uhttp://www.medicis.tv/ par défaut

Référence : http://multimedia.bnf.fr/unimarc_trad/B856-6-2010.pdf

Librairie python UNIMARC/MARC21

Le package pyunimarc contient les principales méthodes pour créer une notice aux formats **Unimarc** et **Marc21**. Ce package doit être importé en début de programme :

```
import pyunimarc
```

Cette librairie permet de créer un fichier de transmission au format Unimarc/Marc21 (*.mrc). Les fichiers générés par cette librairie peuvent être transformé au format *.mrk par un outil comme MarcEdit.

Lancement du programme

Le programme se lance en ligne de commande : `python marc_generator-xx.py`.

```
def main():
    as_unimarc()

if __name__ == '__main__':
    main()
```

Programme de génération des notices

Description des données en entrée, de l'algorithme de génération des notices et des données en sortie.

Données en entrées

Les données en entrée sont extraites de la bases de données Medici et converties au format

JSON. Elles ont alors accessibles par un simple appel REST à l'adresse suivante:

```
url = 'http://www.preprod.medici.tv/api/movie/?specialview=marcexport'
```

Les données sont chargées par le programme et stockées dans la ststructured : `decoded_data`.

```
u = urllib.urlopen(url)
data = u.read()
decoded_data = json.loads(data)
```

Génération des notices

Date de création

Un timestamp global pour la toutes les notices est ensuite crée:

```
from datetime import datetime
date_created = datetime.now()
```

Algorithme

Les notices sont générées par la méthode `as_unimarc()` qui fonctionne de la manière suivante :

1. création du fichier dans lequel sera stockée la notice au format de transmission
2. création d'un objet `Record()` (notice)
3. création des champs `Field()`
4. enregistrement des objets `Field()` dans `Record()`
5. fermeture du fichier

Structure d'une notice

Une notice (*record*) Unimarc ou Marc21 est composée des rubriques suivantes:

1. un leader
2. un directory
3. une liste de champs

Structure d'un champ

Un champ (*field*) au format Unimarc ou Marc 21 est composé des éléments suivants:

1. un tag
2. un tableau d'indicateurs obligatoires ou optionnels en fonction du tag sélectionné
3. un tableau de sous-rubriques ont le contenu dépend du tag

```

field = Field(
    tag = '701',
    indicators = ['0','1'],
    subfields = [
        '$a', '...',
        '$b', '...',
        '$c', '...',
    ]
)

```

Pour créer un champ de contrôle (001-009) il suffit de ne pas renseigner *indicators* ou *subfields* comme dans l'exemple ci-dessous :

```

field = Field(tag='001', data='MEDICI-991')

```

Le tableau ci-dessous détaille les principales étapes de l'algorithme.

DISCLAIMER: Le code en rouge signale un problème d'encodage dans l'extraction de la base de données qui doit être corrigé pour que la génération du champ correspondant puisse fonctionner correctement.

Blocs d'informations codées (100x) - descriptives (200x) - Autres champs uniques

Voir la documentation dans le code.

Informations casting (700x)

Pour remplir les champs 700x l'algorithme itère sur les tableaux "creators", "interpreters", etc... de la structure [decoded_data](#).

Voir la documentation dans le code.

Vérification des notices

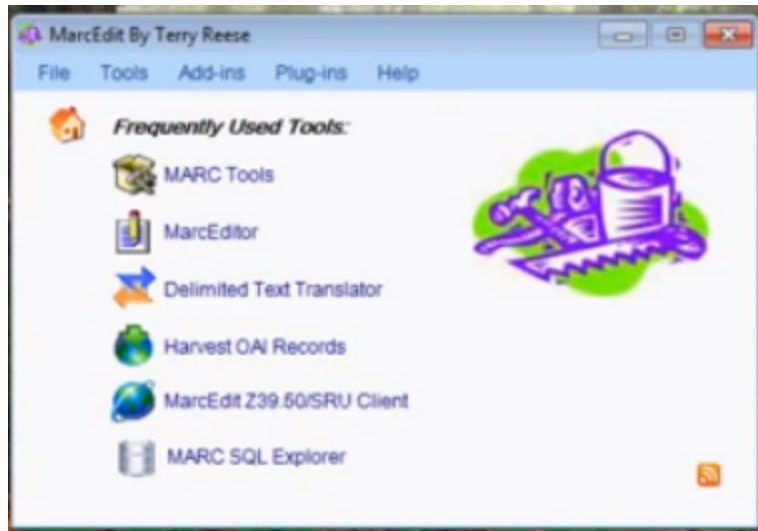
Les notices générées peuvent être vérifier au moyen du programme MarcEdit.

Installation

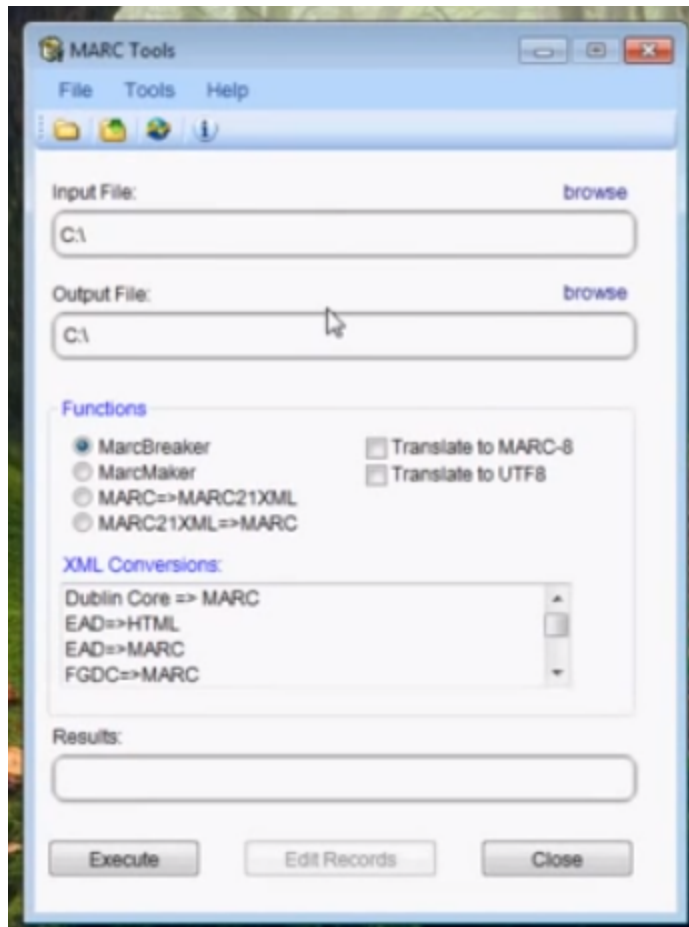
Télécharger le programme d'installation à l'adresse suivante : <http://marcedit.reeset.net/> et lancer le programme d'installation.

Procédure de vérification

Lancer MarcEdit

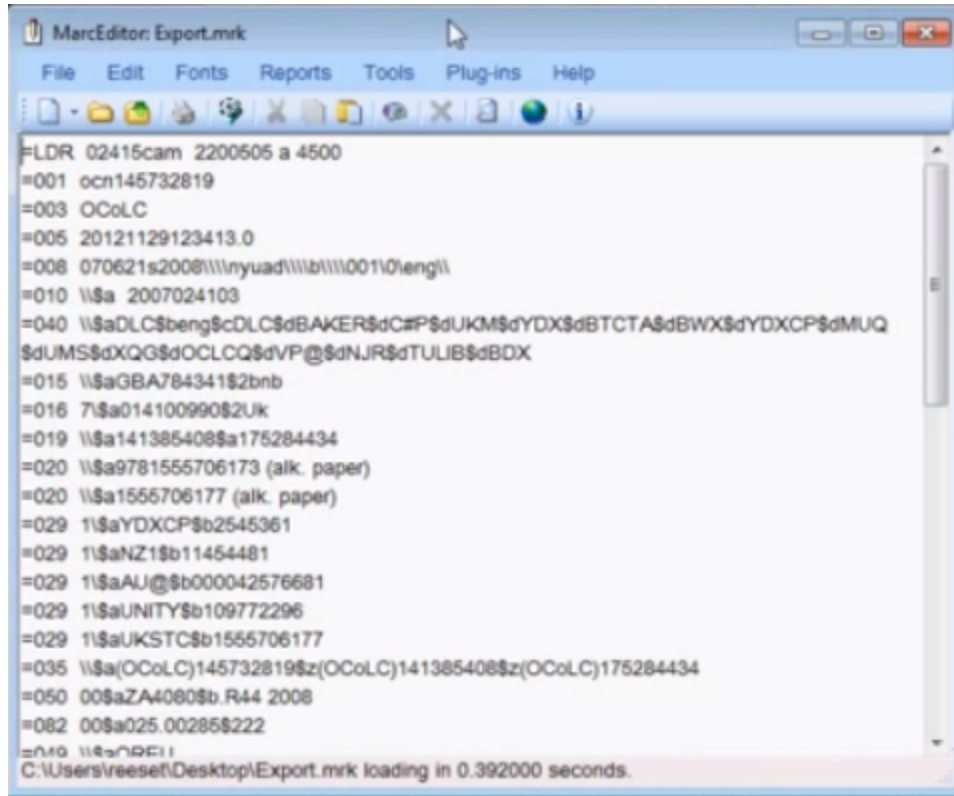


Convertir un fichier *.mrc en fichier *.mrk



Placer le fichier *.mrc dans Input File et spécifier le fichier *.mrk dans Output File puis lancer la commande Execute

Vérifier le contenu du fichier généré



Lancer le programme MarcEditor et vérifier les champs générés.

Un tutoriel est disponible ici : <http://www.youtube.com/watch?v=zP4x-4hcVQ4>.