Bibliostratus : spécifications techniques

Description technique des mécanismes en œuvre dans les différents modules du logiciel **Bibliostratus**

|  |  |
| --- | --- |
| **Version du document** | 1.0 |
| **Version de Bibliostratus décrite** | 1.26 |
| **Date de dernière mise à jour** | 27/03/2019 |

Table des matières

[Module central (blanc) de Bibliostratus : alignements – notices bibliographiques 2](#_Toc4615177)

[Fichier en entrée : format 2](#_Toc4615178)

[[TEX] Monographies imprimées 3](#_Toc4615179)

[[DVD] Documents vidéos (CD, DVD) 3](#_Toc4615180)

[[AUD] Documents sonores (CD, DVD) 3](#_Toc4615181)

[[PER] Périodiques 3](#_Toc4615182)

[[CAR] Cartes 3](#_Toc4615183)

[[PAR] Partitions 4](#_Toc4615184)

[Processus d’alignements 4](#_Toc4615185)

[Erreurs les plus fréquentes 4](#_Toc4615186)

[Alignement par ARK 5](#_Toc4615187)

[Alignement par FRBNF / PPN 5](#_Toc4615188)

[Alignement par EAN 5](#_Toc4615189)

[Alignement par ISBN 6](#_Toc4615190)

[Alignement par n° commercial (CD-DVD) 6](#_Toc4615191)

[Alignement par ISSN (périodiques) 6](#_Toc4615192)

[Alignement par mots (Titre-Auteur-Date-Lieu de publication) 6](#_Toc4615193)

[Mécanismes de contrôle 7](#_Toc4615194)

[Fichier en entrée 8](#_Toc4615195)

[Fichiers en sortie 8](#_Toc4615196)

[Les fichiers de résultats 8](#_Toc4615197)

[Module central (blanc) de Bibliostratus : alignements – notices d’autorité 8](#_Toc4615198)

[Processus d’alignement 9](#_Toc4615199)

[Alignement sur ARK 9](#_Toc4615200)

[Alignement sur FRBNF ou PPN 9](#_Toc4615201)

[Alignement sur ISNI 9](#_Toc4615202)

[Fichiers en sortie 9](#_Toc4615203)

[Modules périphériques de Bibliostratus : fonctionnalités facilitatrices pour l’usager 9](#_Toc4615204)

[Module (bleu) de conversion de fichier de notices en fichiers tabulés 9](#_Toc4615205)

[Extraction de notices bibliographiques - Principes 9](#_Toc4615206)

[Extraction de notices bibliographiques - Zones exportées dans les fichiers en sortie 10](#_Toc4615207)

[Extraction de notices d’autorité – Principes 11](#_Toc4615208)

[Extraction de notices d’autorité - Zones utilisées pour générer chaque élément d’information 11](#_Toc4615209)

[Module (rouge) d’export de notices BnF 12](#_Toc4615210)

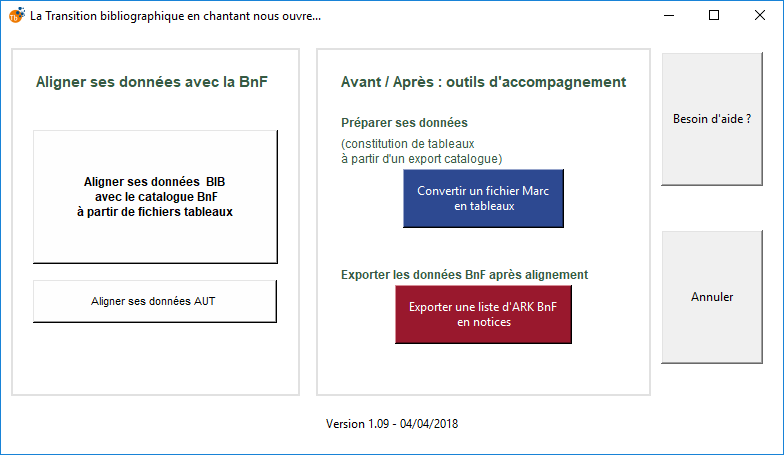
[Option « Fichier d’1 colonne / Fichier à 2 colonnes » 12](#_Toc4615211)

[Option « Données d’autorité liées » 12](#_Toc4615212)

[Option iso2709 / XML 12](#_Toc4615213)

[Option Unimarc / Intermarc 12](#_Toc4615214)

[Préférences 12](#_Toc4615215)



# Module central (blanc) de Bibliostratus : alignements – notices bibliographiques

Ce module prend en entrée un fichier tabulé, encodé en UTF-8, contenant pour chaque notice à aligner les métadonnées qui seront exploitées par le programme pour trouver la notice BnF correspondante.

La préparation d’un tel fichier peut parfois s’avérer plus délicate que prévue, en fonction du SIGB et de la base de données sources. Le logiciel intègre donc des modules facilitateurs décrits dans le chapitre 2 : Modules facilitateurs pour l’usager.

## Fichier en entrée : format

* Fichier à colonnes : une ligne par notice, une colonne par élément d’information.
* Séparateur : tabulation
* Encodage : UTF-8 sans BOM

Selon le type de document, les métadonnées exploitables ne sont pas les mêmes. Le programme distingue donc plusieurs types de documents, et il n’est pas possible de mettre dans un même fichier des livres et des DVD, ou des livres et des périodiques

### [TEX] Monographies imprimées

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**ISBN**

**EAN**

Titre

Auteurs

Date

Tome

Editeur

### [DVD] Documents vidéos (CD, DVD)

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**EAN**

**Numéro commercial**

Titre

Auteur

Date

Editeur

### [AUD] Documents sonores (CD, DVD)

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**EAN**

**Numéro commercial**

Titre

Auteur

Date

Editeur

### [PER] Périodiques

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**ISSN**

Titre

Auteur

Date

Lieu de publication

### [CAR] Cartes

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**ISBN**

**EAN**

Titre

Auteur

Date

Editeur

Echelle

### [PAR] Partitions

Numéro de notice

**FRBNF (préfixé FRBN ou FRBNF) / PPN (préfixé PPN) initial**

**ARK initial**

**EAN**

**Référence commerciale**

Titre

Titre de partie

Auteur

Date

Editeur

## Processus d’alignements

Pour chacun des types d’alignements choisis, le programme essaie d’aligner :

* D’abord sur les identifiants (éléments en gras) dans l’ordre de préférence suivant :  
  (il ne passe au suivant que si le précédent n’a rien donné)[[1]](#footnote-1) :
  + ARK
  + FRBNF ou PPN
  + EAN
  + ISBN
  + Numéro commercial
  + ISSN

A défaut, c’est une recherche par mot-clé qui s’effectue (Titre-Auteur-Date)

Lors de la recherche par identifiants, sauf pour les ARK, des contrôles s’effectuent entre les métadonnées en entrée et celles trouvées dans les notices.

Ces contrôles sont détaillés ci-dessous.

### Erreurs les plus fréquentes

**Rappel : le programme est tributaire de la qualité des données contenues dans le fichier fourni, et des données du catalogue de la BnF.** Des erreurs dans la manière d’alimenter les zones ou les colonnes génèreront des erreurs dans la réponse du programme.

* Les colonnes attendues ne correspondent pas :
  + Il manque une colonne dans le fichier en entrée, si bien que le programme utilise la colonne Titre comme s’il s’agissait d’un EAN, par exemple
  + Oubli de sélectionner le bon type de document pour le fichier mis en entrée  
    
* Les zones codées Type de document / Type de contenu ont été mal renseignées  
  🡪 un DVD de film se retrouve dans le fichier des imprimés, et l’alignement proposé pointe sur le roman qui a inspiré le film…
* L’encodage du tableur n’est pas de l’UTF8 sans BOM, ce qui fait planter le programme
* Les données sont passées par Excel, qui a transformé les ISBN ou EAN en données du type « 9,78E15711 ».  
  Idem sur les colonnes Dates, notamment

### Alignement par ARK

Recherche sur l’ARK. Quelle que soit la notice trouvée, son ARK (identique ou différent à l’ARK en entrée) est retourné.

### Alignement par FRBNF / PPN

La même colonne permet de requêter soit par FRBNF (numéro BnF), soit par PPN (numéro Sudoc).

Dans les 2 cas, si la notice Sudoc contient le numéro BnF, on peut récupérer soit l’identifiant Sudoc soit l’identifiant BnF.

#### Recherche par FRBNF

1. Recherche sur le nombre complet avec préfixe sur le critère de recherche bib.otherid du SRU BnF.
2. Si pas de résultat, recherche des 8 premiers chiffres comme numéro de notice actuel + contrôle sur ISBN
   * Si le contrôle sur ISBN échoue (ou si pas d’ISBN), contrôle sur le titre  
     Si le contrôle sur le titre est valide et qu’il y a un numéro de volume, contrôle sur le numéro de volume
3. Si pas de résultat, recherche sur les 9 premiers chiffres comme ancien numéro de notice (zone 9XX) + contrôle sur ISBN
   * Si le contrôle sur ISBN échoue (ou si pas d’ISBN), contrôle sur le titre + volume
4. Si pas de résultat, recherche sur les 8 premiers chiffres comme ancien numéro de notice (zone 9XX) + contrôle sur ISBN
   * Si le contrôle sur ISBN échoue (ou si pas d’ISBN), contrôle sur le titre + volume

#### Recherche par PPN

Si le préfixe dans cette colonne n’est pas « FRBN » ou « FRBNF », mais « PPN », Bibliostratus ouvre la notice Sudoc pour voir s’il y a un FRBNF ou un ARK proposé.

### Alignement par EAN

* Recherche par EAN dans le catalogue BnF (SRU : critère bib.ean) ou dans le Sudoc (API ean2ppn)
* contrôle sur titre, volumaison et date
* Si échec, recherche dans toute la notice
* Si échec, on cherche l’EAN sans faire de contrôle sur les titre, volumaison et date  
  Si au moins une notice est trouvée, son identifiant est récupéré, mais avec en colonne « Méthode » un message mentionnant un « Problème » lors du contrôle de conformité des métadonnées

### Alignement par ISBN

* Recherche par ISBN dans le catalogue BnF (SRU : critère bib.isbn) ou dans le Sudoc (API isbn2ppn)
* Contrôle sur le titre + numéro de volume
* Si échec du contrôle : recherche ISBN + Auteur (mot d’auteur le plus long dans la zone Auteur du fichier en entrée)   
  + contrôle sur la date
* Si échec : recherche sur l’ISBN nettoyé : uniquement le premier ISBN (s’il y a en entrée plusieurs ISBN séparés par des ponctuation), en supprimant les espaces, les tirets, etc.  
  + contrôle sur le titre et la date
* Si échec : recherche sur ISBN nettoyé + Auteur  
  + contrôle sur la date
* Si échec, conversion de l’ISBN en entrée : ISBN10 en ISBN13 ou ISBN13 en ISBN10  
  + mêmes contrôles
* Si échec, on refait les mêmes recherches, mais dans toute la notice (critère de recherche SRU bib.anywhere)

Si la notice initiale contenait des métadonnées Titre ou Auteur, et que l’alignement sur ISBN a eu un échec sur le contrôle des métadonnées des notices trouvées (titre non conforme), un message « Problème » est ajouté en colonne Méthode d’alignement.

### Alignement par n° commercial (CD-DVD)

Numéro commercial : 071$b (source) + 071$a (identifiant interne à la source)

Recherche du numéro commercial dans le critère bib.comref :

* Si le n° commercial a plusieurs mots :
  + Le dernier mot est reconnu comme identifiant
  + Ceux qui le précèdent sont la source  
    Construction d’une requête : bib.anywhere all <source> and bib.comref all <identifiant>
* S’il y a un seul mot :   
  Requête bib.comref all <identifiant>

+ Contrôles

* Contrôle sur la 071$a de la notice BnF : si le numéro commercial en entrée est présent dans la 071$a de la notice BnF, ou si la notice BnF contient le numéro commercial initial
* + contrôle sur titre, volumaison et date

Le Sudoc ne propose pas de web service sur ce critère.

### Alignement par ISSN (périodiques)

* Recherche de l’ISSN dans le critère de recherche SRU bib.issn  
  + contrôle sur le type de notice pour chaque notice trouvée
* Si échec : recherche dans le Sudoc (web service issn2ppn)

### Alignement par mots (Titre-Auteur-Date-Lieu de publication)

Titre nettoyé (suppression des accents et ponctuations, tout passé en minuscules) + conservation uniquement des mots des plus de 1 lettre

Cas particuliers :

* Périodiques : Recherche sur Titre + Auteur + Date + Lieu de publication  
  Contrôle sur le titre  
  Contrôle sur la date  
  Si aucune réponse 🡪 on modifie la date à -3 ans et +3 ans
* Cartes : Recherche sur Titre + Auteur + Date + Echelle
* Partitions : Recherche sur Titre + complément de titre + Auteur + Date  
  Contrôle sur les sous-zones de titre principal ou titre d’ensemble titre  
  Contrôle sur les sous-zones de titre principal ou de titre de partie pour le complément de titre

Le Sudoc ne propose pas de web service sur ces critères. Bibliostratus lance donc une « recherche avancée » dans l’interface web et navigue dans les pages de résultats HTML. Cette fonction peut être chronophage : c’est pourquoi elle peut être désactivée (case à décocher) dans le formulaire de recherche.

### Mécanismes de contrôle

#### Sur l’ISBN

Récupération et nettoyage de la racine de l’ISBN (9 premiers chiffres d’un ISBN10, ou 4e-12e chiffres d’un ISBN13) du fichier en entrée

Recherche de l’ISBN dans la notice BnF (010$a, première occurrence uniquement) et nettoyage de sa racine.

Comparaison des 2 valeurs

#### Sur le titre

Le logiciel reprend d’abord la chaîne de caractères Titre en entrée et la normalise :

* Suppression des ponctuations et des espaces
* Suppression des diacritiques
* Passage de toutes les lettres en minuscules

Il compare la chaîne de caractères obtenue au titre de la notice BnF, cherchée successivement en :

* 200$a (titre)
* Si échec, successivement : 200$e (complément de titre), 200$i (titre de partie), 750$a (autre forme de titre), 753$a (autre forme de titre de périodique), 500$a (titre uniforme), 500$e (titre uniforme, complément), 503$a, 503$e (titre de forme), 540$a, 540$e (titre ajouté par le catalogueur), 410$a (collection), 225$a (collection), 461$t (titre d’ensemble), 464$t (titre de partie)
* Si échec, il fait la comparaison sur la moitié de la longueur du titre

Pour les partitions, le « titre de partie » est contrôlé dans les zones : 464$t, 200$a+200$e, 225$e

#### Sur la date

Récupération et nettoyage de la date du fichier en entrée (on ne conserve que les chiffres)

Récupération de la date de la notice BnF :

* 4 premiers caractères de la zone 100$a -> on prend aussi les années n-1 et n+1
* Zone 210$d : on ne garde que les chiffres

Si la date en entrée est présente dans les dates récupérées de la notice BnF, le contrôle est valide.

#### Sur le numéro de tome/volume

Tous les chiffres romains sont passés en chiffres arabes pour faciliter les comparaisons.

Récupération des chiffres présents dans les zones :

* 200$h
* 200$u
* 461$v
* (si ces zones n’existent pas, récupération des chiffres en 200$a)

Si le numéro de tome en entrée est présent dans un des numéros trouvés dans la notice BnF, le contrôle est valide.

## Fichier en entrée

Fichier tabulé, encodé en UTF-8, dont les colonnes correspondent à celles de l’option choisie (TEX, DVD, AUD, PER, CAR, PAR)

## Fichiers en sortie

Le fichier renvoie :

* Un ou plusieurs fichiers donnant le résultat des alignements (selon l’option 1 fichier / plusieurs fichiers choisie dans le formulaire)
* Un fichier « statistiques » qui renvoie global de notices pour lequel 1 ARK a été trouvé ; ou 2 ARK ; ou 3 ARK, etc.  
  Ce fichier contient également (si besoin) la liste des problèmes rencontrés : indisponibilité d’un web service pendant une certaine durée, ou URL renvoyant ponctuellement un message d’erreur

### Les fichiers de résultats

Ces fichiers contiennent systématiquement les colonnes suivantes :

* Le numéro de notice en entrée
* Le nombre d’identifiants trouvés
* La liste des identifiants trouvés (ARK ou PPN, séparés par une virgule)
* La méthode qui a permis de trouvé le(s) identifiant(s)
* Toutes les métadonnées fournies en entrée
* Si la case « Métadonnées sommaires » a été cochée dans le formulaire : les métadonnées BnF de l’ARK ou des ARK trouvés :
  + Titre = 200$a + 200$e
  + Nom du premier auteur = 700$a ou 710$a
  + Prénom du premier auteur = 700$b ou 710$b
  + Tous auteurs = 200$f
  + Date = 210$d

# Module central (blanc) de Bibliostratus : alignements – notices d’autorité

Ce module permet d’aligner des notices d’autorité avec les autorités BnF – uniquement des personnes physiques et organisations.

Deux types de fichier en entrée sont possibles :

* Soit on a exporté un fichier d’autorités : les métadonnées servant à l’alignement sont donc les points d’accès (nom + dates)
* Soit on a exporté un ensemble de notices bibliographiques : le logiciel trouvera d’abord le document, pour remonter ensuite à l’auteur

Ce module a été beaucoup moins développé que le module d’alignement sur les données bibliographiques.

## Processus d’alignement

Comme pour les notices bibliographiques : alignement d’abord sur les n° internes BnF (ARK ou FBRNF/PPN).

À défaut (si absents ou aucun résultat trouvé), utilisation de la colonne ISNI

À défaut (si ISNI absent ou aucun résultat trouvé), utilisation du point d’accès (combinaison Nom – Prénom – Date de début)

* S’il n’y a pas de date de début mais une date de fin, l’alignement se fait sur :
  + point d’accès contient l’expression Nom prénom
  + + point d’accès contient date de fin
  + + contrôle : date de fin de la notice trouvée = date de fin fournie en entrée

À défaut : recherche du point d’accès dans la base [isni.org](http://isni.org/)

Toutes les colonnes indiquées dans le formulaire doivent être présentes, même si elles ne sont jamais renseignées.

### Alignement sur ARK

Le contenu de la colonne ARK est nettoyé : seuls les caractères qui suivent la mention ark:/12148/+9 caractères sont conservés, le reste est nettoyé.

Si la notice à laquelle correspond l’ARK BnF a été fusionnée entre temps, le nouvel ARK est renvoyé pour l’alignement.

### Alignement sur FRBNF ou PPN

Le tableau est censé contenir une seule valeur. S’il y en a plusieurs, seule la première est prise en compte (séparateurs : ponctuation ou espace). Soit la valeur commence par FRBNF ou FRBN et c’est un alignement par FRBNF ; soit la valeur commence par PPN et c’est un alignement sur IdRef.

Les autres valeurs sont ignorées.

### Alignement sur ISNI

L’ISNI est une valeur sur 16 chiffres, préfixée ou non « isni.org ». Tout autre valeur est ignorée.

## Fichiers en sortie

Cf. la description des fichiers en sortie pour les alignements bibliographiques

# Modules périphériques de Bibliostratus : fonctionnalités facilitatrices pour l’usager

## Module (bleu) de conversion de fichier de notices en fichiers tabulés

### Extraction de notices bibliographiques - Principes

Chacun peut directement le module d’alignement, en alimentant chaque colonne selon les sources d’information qu’il estime pertinentes (en fonction de ce qu’il connaît de ses données et de ses objectifs).

Mais pour faciliter cet alignement, le logiciel propose un module qui, à partir d’un fichier exporté d’un catalogue au format Unimarc (iso2709 ou XML), propose des tableaux à utiliser dans le module d’alignements

Pour chaque notice, extraction

* Du type de notice (zone codée)
* Du type de document (zone codée)
* De différentes métadonnées définies ci-dessous :
  + Numéro FRBNF : 035$a
  + Numéro ARK BnF : 033$a
  + Numéro ISBN : 010$a
  + Numéro EAN : 038$a
  + Numéro ISSN : 011$a
  + Numéro commercial éditeur (documents sonores et vidéos, partitions) : 073$a
  + Titre : 200$a + 200$e  
    Pour les périodiques : on privilégie le titre clé quand il existe : 530$a
  + Auteur : concaténation des zones suivantes : 700$a, 700$b, 710$a, 710$b, 701$a, 701$b, 711$a, 711$b, 702$a, 702$b, 712$a, 712$b

À défaut : 200$f

* + Éditeur : 210$c
  + Lieu de publication : 210$a
  + Date de publication : zone codée 100$a ; ou à défaut 210$d ; ou à défaut 219$d, 219$i, 219$p
  + Titre de partie : 464$t

À défaut : 200$a, 200$e

* + Numéro de volume : 200$h
  + Échelle : 123$a

Puis, selon le type de notice/type de document, certaines seulement de ces métadonnées seront conservées dans le fichier en sortie, pour être utilisés comme critères d’alignements

### Extraction de notices bibliographiques - Zones exportées dans les fichiers en sortie

#### Monographies – texte imprimé

FRBNF / PPN

ARK

ISBN

EAN

Titre

Auteurs

Date de publication

Numéro de tome

Éditeur

#### Monographies – son, vidéo

FRBNF / PPN

ARK

EAN

ID commercial audio

Titre

Auteurs

Date de publication

Éditeur

#### Périodiques

FRBNF / PPN

ARK

ISSN

Titre clé, ou à défaut Titre

Auteurs

Date de publication

Lieu de publication

#### Cartes

FRBNF / PPN

ARK

ISBN

EAN

Titre

Auteurs

Date

Éditeur

Échelle

#### Partitions

FRBNF / PPN

ARK

EAN

Référence commerciale

Titre

Titre de partie

Auteurs

Date

Éditeur

#### Autres types de notice-type de document (pas d’alignement programmé)

FRBNF / PPN

ARK

EAN

Titre

Auteurs

Date de publication

### Extraction de notices d’autorité – Principes

Les principes sont les mêmes que pour les notices bibliographiques : on met en entrée un fichier Unimarc de notices d’autorité, et le programme répartit ces notices par types (personnes physiques, organisations, mots-matières, oeuvres)

### Extraction de notices d’autorité - Zones utilisées pour générer chaque élément d’information

* + ARK : 033$a
  + FRBNF / PPN : 035$a
  + ISNI : 010$a
  + Nom : 200$a

À défaut (dans le cas des ORG) : 210$a

* + Complément du nom : 200$b

À défaut (dans le cas des ORG) : 210$b

* + Date de début : 103$a  
    Si non renseigné 🡪 200$f
* avant le "–" s’il y en a un
* sinon toute la zone
  + Date de fin : 103$b  
    Si non renseigné 🡪 200$f :
* après le tiret uniquement
* sinon, aucune valeur récupérée

## Module (rouge) d’export de notices BnF

À partir d’un fichier donnant une liste d’ARK (un ARK par ligne), ce module permet d’extraire des notices BnF en interrogeant le SRU de la BnF.

La liste des ARK peut contenir des ARK de notices bibliographiques, ou des ARK de notices BnF (mais pas mélanger les 2).

### Option « Fichier d’1 colonne / Fichier à 2 colonnes »

Il est possible de mettre en entrée 2 types de fichiers au choix :

* Un fichier à une colonne, contenant une liste d’ARK et/ou de PPN (préfixés « PPN »)  
  En ce cas les notices sont exportées telles que dans leur base source.
* Un fichier à deux colonnes : N° de notice | ARK ou PPN  
  En ce cas, chaque notice exportée est réécrite : le contenu de la zone 001 de la notice initiale, qui contient le numéro BNF ou Sudoc, est déplacé en 035. A la place, le numéro de notice « local » est écrit en 001.  
  De même, la zone 003 est déplacée en 033.  
  En revanche les liens aux autres notices (notamment Autorités) ne sont pas réécrits.

### Option « Données d’autorité liées »

Si l’utilisateur met en entrée une liste d’ARK de notices bibliographiques, et demande à récupérer aussi le fichier de notices d’autorité liées, le programme ouvre chaque notice bibliographique et regarde les zones suivantes :

100, 141, 143, 144, 145, 600, 603, 606, 607, 609, 610, 616, 617, 700, 702, 703, 709, 710, 712, 713, 719, 731

Il identifie le $3 (qui fournit l’identifiant de la notice d’autorité) et va chercher la notice correspondante pour l’agréger dans le fichier d’autorités en sortie.

Une même autorité n’est exportée qu’une seule fois.

### Option iso2709 / XML

Possibilité que les fichiers en sortie soient en iso2709 (encodés en UTF-8) ou en XML (encodés en UTF-8 aussi)

### Option Unimarc / Intermarc

Possibilité de choisir le format MARC en sortie.

# Préférences

Depuis la version 1.26, il existe un fichier preferences.json (dans le dossier main/files) qui permet d’enregistrer plusieurs paramètres entre deux sessions.

Ce fichier peut être édité via le menu principal



Il permet de personnaliser les paramètres suivants :

* un proxy réseau http et https, pour que Bibliostratus puisse l’utiliser et accéder à internet en cas de réseau institutionnel bridé
* pour le module bleu (marc > tableaux), on peut indiquer que le format en entrée des fichiers est du Marc21 et non de l’Unimarc. Ce seront alors les zones de titre, ISBN, Auteur, etc. spécifiques au Marc qui seront récupérées lors de l’extraction des métadonnées pour générer les tableurs.
* une option pour activer l’horodatage des alignements : si on constate un ralentissement dans les alignements dans le cas de gros fichiers, activer cette fonction permet d’obtenir pour chaque ligne en entrée une information en sortie sur l’heure (et minute et seconde) exacte de l’alignement, pour pouvoir vérifier si Bibliostratus ralentit progressivement dans le cas de gros fichiers à traiter
* une option pour déclarer (dans un fichier à part) une liste de mots vides, à ne pas prendre en compte lors d’alignements. Si l’un de ces mots est présent dans la zone de titre du document à aligné, il n’est pas utilisé lors des alignements par Titre-Auteur-Date

En cas d’erreur de saisie, on peut rétablir les valeurs par défaut du logiciel

# Bibliostratus en ligne de commande

Le fichier **bibliostratus-cli.py** permet d’utiliser les fonctions d’extraction et d’alignement de bibliostratus, pour les 4 modules.

* Alignement bibliographique (module blanc) : **--action bib2id**
* Alignement autorités (module blanc) : **--action aut2id**
* Conversion de fichiers marc en tableaux (module rouge) : **--action marc2tables**
* Extraction de notices Marc à partir d’une liste d’identifiants ARK ou PPN :   
  **--action ark2records**

A par le nom du fichier en entrée, tous les paramètres sont optionnels : ils prennent alors les mêmes valeurs par défaut que celles du formulaire

Exemples

**bib2id** (utilisant les valeurs par défaut pour plusieurs arguments)

python bibliostratus-cli.py --action bib2id --file main\examples\mon\_impr.tsv --prefix cli\_test

**aut2id**

python bibliostratus-cli.py --action aut2id --file main\examples\aut\_align\_aut.tsv

**marc2tables**

python bibliostratus-cli.py --action marc2tables --file main\examples\3bibrecords.xml --filetype xml-utf8 --recordtype bib --prefix cli\_test

**ark2records**

* Export de notices BIB en XML + AUT liées

python bibliostratus-cli.py --action ark2records --file main\examples\listeARKbib.tsv --recordtype bib --colnum 1 --get\_authors oui --output\_format unimarc --output\_file xml --encoding utf8 --dir main --prefix cli\_test

* Export tabulé de BIB (zones 200 et 700)

python bibliostratus-cli.py --action ark2records --file main\examples\listeARKbib.tsv --recordtype bib --colnum 1 --output\_format unimarc --output\_file tab --fields 200$a;700 --dir main --prefix cli\_test

## Paramètres communs

### --file : Nom du fichier en entrée

Obligatoire

Texte libre

### --dir : Dossier de destination

Texte libre

### --prefix : Préfixe pour le(s) fichier(s) en sortie

Texte libre

## --action bib2id

### --doctype : Type de documents à aligner

Valeurs autorisées : tex, vid, aud, per, car, par,

Défaut : tex

### --align\_pref : Préférences d'alignement (bnf / abes)

Valeurs autorisées : bnf, abes, sudoc

Défaut : bnf

### --sudoc\_kw : Recherche par mots-clés dans le Sudoc

Valeurs autorisées : oui, non

Défaut=non

### --headers : Fichier avec en-têtes

Valeurs autorisées : oui, non

Défaut=oui

### --files\_nb : Nombre de fichiers en sortie

Valeurs autorisées : 1, 3

Défaut=1

### --metas : Récupérer les métadonnées simples

Valeurs autorisées : oui, non

Défaut =non

## --action aut2id

### --doctype : Type de notices à aligner

Valeurs autorisées : pers, org, ram, rameau

Défaut : pers

### --align\_pref : Préférences d'alignement (bnf / abes)

Valeurs autorisées : bnf, abes, idref

Défaut : bnf

### --isni : Relancer la recherche dans isni.org

Valeurs autorisées : oui, non

Défaut=non

### --headers : Fichier avec en-têtes

Valeurs autorisées : oui, non

Défaut=oui

### --files\_nb : Nombre de fichiers en sortie

Valeurs autorisées : 1, 3

Défaut=1

## --action marc2tables

### --filetype : Type de fichier en entrée

Valeurs autorisées : iso2709-utf8, iso2709-8859-1, xml, xml-utf8

Défaut : iso2709-utf8

### --recordtype : Type de notices

Valeurs autorisées : bib, aut, bib2aut

Défaut : bib

## --action ark2records

### --recordtype : Type de notices

Défaut : bib

Valeurs autorisées : bib, aut

### --colnum : Nombre de colonnes en entrée [1 / 2]

Défaut=1

Valeurs autorisées : 1, 2

### --headers : Fichier avec en-têtes

Défaut : oui

### --get\_authors : Récupérer les auteurs liés

Défaut : non

### --get\_subjects : Récupérer les notices Sujet liées

Défaut : non

### --get\_works : Récupérer les notices Oeuvres liées

Défaut : non

### --output\_format : Format en sortie des notices

Valeurs autorisées : unimarc, intermarc, unimarc-anl, intermarc-anl

Défaut : unimarc

### --output\_file : Format en sortie du fichier

Valeurs autorisées : iso2709, xml, tab

Défaut : iso2709

### --encoding : Encodage en sortie

Valeurs autorisées : utf8, utf-8, iso-8859-1, iso88591, 8859-1

Défaut : utf8

### --fields : Zones spécifiques à récupérer

Zone de texte libre

1. Chaque test ne s’effectue que si l’information existe bien pour le type de document traité. [↑](#footnote-ref-1)