

À partir d'une liste de titre et leur localisation extraite de notre SGB nous voudrions nous assurer que chaque localisation existe bien dans le SUDOC.

Pour réaliser cette opération, nous allons utiliser [le web service multiwhere](#). Ce web service prend en paramètre un ou plusieurs PPN et fournie en réponse la liste des RCR localisés sous la notice.

Nous allons ainsi, pour chaque titre présent dans notre fichier appeler le web service de l'ABES pour obtenir la liste des rcr localisés sous la notice. Nous allons ensuite extraire chaque bibliothèque Alma présente dans la colonne *Disponibilité* pour la transformer en RCR. Nous nous assurerons ensuite que le RCR de chaque localisation est bien présent dans cette liste retournée par le SUDOC.

## 1 Construction des urls et obtention de la liste des rcr localisés sous la notice

## 1.1 Import du fichier dans Open Refine

1.1.1 Téléchargez le fichier [fichier\\_source\\_cas\\_pratique\\_2.xlsx](#). 1.1.2 Ouvrez Open Refine 1.1.3 Créer un projet (1.)> Récupérer les données à partir de cet ordinateur (2.) > Browse (3.) > Suivant >>(4.)

1 / 7

Si nécessaire, ajustez les options d'importation (1.) Renommez votre projet (2.) Créez le projet (3.)

**OpenRefine** Un outil puissant pour travailler avec des données désordonnées. Nouvelle version [Télécharger OpenRefine v3.7.7 maintenant.](#)

Créer un projet Recommencer Configurer les options pour l'analyse syntaxique 2. Nom du projet: web service.xlsx 3. Créer un projet >

Ouvrir un projet  
Importer un projet  
Langue

Créateur / Éditeur	Titre	Date de modification	Langue	Date de création	ISBN	Numéro de notice	Disponibilité	Identifiant MMS	underReview	contributionEditor	resultFromNetwork	Identifiant de handle	statut de préservation
1. Livre (Livre - Physique) Par Sordello, Jérôme (Paris : Amphora : C 2015)	La bible du running : route et trail running / Jérôme Sordello	2023-10-10T12:59:23Z	Français	2018-09-17T10:01:54Z	978-2-85180-925-4	(PPN)188711430	Disponibilité:Version physique à Bib. STAPS(33522105): APSA: TA 152; TA 152 (1 exemplaire, 1 disponible(s))	991001184449704672					
2. Livre (Livre - Physique) Par Adam, Guillaume : cours de marathon et ultra-trail (Paris : Leduc Editions, DL 2023)	Ma bible du running, du trail et du marathon / Guillaume Adam	2023-10-10T15:08:38Z	Français	2023-04-20T15:03:32Z	978-10-285-2786-0	(PPN)269032126	Disponibilité:Version physique à Bib. STAPS(33522105): APSA: TA 181; TA 181 (1 exemplaire, 1 disponible(s))	99100595158504672					
3. Livre (Livre - Physique) Par Ratsimihah, Benjamin (Paris : Larousse, DL 2021)	Trail-running : le guide complet pour se lancer et progresser // Benjamin Ratsimihah et Florence Heimburger	2023-10-10T14:53:59Z	Français	2021-05-11T14:04:00Z	978-2-03-597145-6	(PPN)254475183	Disponibilité:Version physique à Bib. STAPS(33522105): APSA: TA 172; TA 172 (1 exemplaire, 1 disponible(s))	991005810468904672					
4. Livre (Livre - Physique) Par Bazin, Sylvain : journaliste spécialisé dans la randonnée (Paris : Amphora : C 2015)	Trail running : préparez vos défis : des courses nature à l'ultra-trail / Sylvain Bazin, Jean-Marc Delorme	2023-10-10T11:23:43Z	Français	2018-09-17T09:57:53Z	978-2-85180-918-6	(PPN)18625430X	Disponibilité:Version physique à Bib. de FIUT(331922101): Biologie. Santé - 1 - Grande salle, QT 260.5 A (1 exemplaire, 1 disponible(s))	991000891489704672					

Considérer les données comme 1.

**Fichiers Excel**  
Fichiers JSON  
Fichiers texte à base de lignes  
Fichiers CSV / TSV / séparateur  
Fichiers texte à largeur de champ fixe  
PC-Axis text files  
Fichiers MARC  
Fichiers JSON-LD  
Fichiers RDF/N3

Version 3.5.2 [e36f64e]  
Préférences  
Aide  
À propos

Feuilles à importer  
☒ Tout sélectionner ☐ Tout désélectionner  
web service.xlsx#results 9 rows

☐ Ignorer la ou les première(s) ligne(s) du début du fichier  
☒ Analyser la ou les ligne(s) suivante(s) comme des entêtes de colonnes  
☐ Ignorer la ou les première(s) ligne(s) de données  
☐ Changer au plus première(s) ligne(s) de données

☒ Conserver les lignes vides  
☒ Enregistrer les cellules vides comme des valeurs nulles  
☐ Indiquer la source du fichier  
☐ stocker le fichier d'archive


Mettre à jour l'aperçu

## 1.2 Nettoyage du PPN

Regardez le contenu de la colonne *Numéro de notice*. Notre PPN est préfixé par la chaîne de caractère "(PPN)". Nous allons devoir nettoyer notre champ sinon le web service de l'ABES ne reconnaitra pas notre identifiant. Pour ce faire, il existe plusieurs méthodes. En voici quelques-unes.

### 1.2.1 En utilisant la fonction remplacer

Sous la colonne *Numéro de notice*, allez dans **Editer les cellules** et cliquez sur **Remplacer**.

Ecran Remplacer Remplacez simplement la chaîne "(PPN)" par une chaîne vide. Laissez vide le champ Remplacer par :

### 1.2.3 En transformant les données

Sous la colonne *Numéro de notice*, allez dans **Editer les cellules** et cliquez sur **Transformer...**

Ecran Remplacer Nous pouvons utiliser plusieurs formules pour nettoyer notre PPN.



**Astuce :** Aidez-vous de l'aperçu pour voir en temps réel l'effet de votre formule.

### Substring

`value.substring(5,14)` Extrait les caractères de la chaîne *value* entre les positions 5 et 14 (Attention, la numérotation commence à 0). Si le deuxième paramètre est absent, extrait les éléments de jusqu'au dernier. Dans notre cas `value.substring(5)` permet d'obtenir le même résultat.

### Replace

`value.replace("(PPN)","")` Dans la chaîne *value*, remplace la chaîne "(PPN)" par la chaîne vide. Il est possible d'utiliser une expression régulière. Elle doit être encadrée par `/`. Exemple :

`value.replace(/^. {5}/, "")` --> Remplace les 5 premiers caractères de la chaîne par une chaîne vide.

### Slice

La chaîne de caractère est considérée comme une liste de caractère. Ainsi on peut exprimer la valeur de notre colonne ainsi : [(,P,P,N,),1,8,6,2,5, 4, 3, 0, X]. Chaque caractère va avoir une position dans la liste.

```
[ ( , P , P , N , ) , 1 , 8 , 6 , 2 , 5 , 4 , 3 , 0 , X ]
0 1 2 3 4 5 6 7 8 9 10 11 12 13
```

Vi la formule, `value[5,14]`, on extrait les valeurs de la position 5 à 14. L'extraction s'arrête juste avant la 14ème position.

### 1.3 Construction de l'URL et appel du web service

Nou allons maintenant construire l'URL pour chaque PPN et appeler le web service multiwhere. Regardons la [documentation du web service multiwhere](#). L'adresse doit être construite ainsi :

```
https://www.sudoc.fr/services/multiwhere/[PPN]&format=text/json
```

Le dernier paramètre de l'URL (*format*) indique que nous voulons recevoir les résultats [au format json](#).

💡 La plupart des webservice fournissent les données au format xml ou json. Dans Open Refine nous vous conseillons de privilégier le deuxième format beaucoup plus facile à manipuler.

#### 1.3.1 À partir de la colonne *numéro de notice*,

cliquez sur `Editer la colonne` > `Ajouter la colonne en moissonnant des urls`. Un nouvel écran apparaît.

The screenshot shows the OpenRefine interface with the 'Ajouter une colonne en moissonnant les données depuis les URL' dialog box open. The dialog has several fields and options:

- Nouveau nom de colonne:** `webservice` (labeled 4).
- Délai de récupération:** `500` millisecondes (labeled 2).
- En cas d'erreur:** `vider la cellule` (labeled 3) and `conserver l'erreur` (labeled 3).
- Entêtes HTTP à utiliser pour la récupération d'URLs:** `Afficher`.
- Indiquer les URL à moissonner:** `https://www.sudoc.fr/services/multiwhere/'+value+'&format=text/json'` (labeled 1).
- Langue:** `General Refine Expression Language (GREL)`.
- Aperçu:** A preview table showing the harvested data with columns 'row' and 'value'.


The preview table shows the following data:


row	value
1.	188711430
2.	269032126
3.	254475183
4.	18625430X
5.	233849424
6.	234198524

#### 1.3.2 Dans la zone de saisie de l'expression (1.),

construisez l'url en concaténant le PPN présent dans la colonne *numéro de notice* (obtenu via le paramètre *value*) avec les autres éléments de l'URL.

```
"https://www.sudoc.fr/services/multiwhere/'+value+'&format=text/json"
```


 **Rappel** Les chaînes de caractères doivent être encadrées par des ' ou des "

 **Rappel** le paramètre **value** permet d'obtenir les valeurs des cellules de la colonne en cours de transformation. Pour appeler les valeurs des cellules des colonnes voisines, utilisez l'expression suivante **cells[nom de la colonne].value**

### 1.3.3 Modifiez le délai de récupération (2.) à 500 millisecondes

### 1.3.4 En cas d'erreur, choisissez de *conserver l'erreur* (4.)

### 1.3.5 Nommez la colonne *réponse web service* (4.) et validez (5.).

 **Attention** : Pour chaque ligne du projet, Open refine va appeler le web service de l'ABES. Ce traitement peut prendre plus ou moins de temps selon la taille du fichier à traiter.

## 2. Extraction des localisations Alma et transformation en RCR

### 2.1 Envoyer chaque localisation sur une ligne unique

Si vous regardez la cellule disponibilité vous remarquerez que chaque mention de disponibilité est séparée par un saut de ligne.

 **Rappel** : En expression régulière le saut de ligne est représenté ainsi `\n`

#### 2.1.1 S'assurer qu'il n'y ait pas de ligne vide

À partir de la colonne *Toutes*, allez dans `Facette > Afficher les valeurs vides par colonne`.

On remarque toutes les colonnes ayant des valeurs vides sont vides pour la totalité de nos entrées. Nous n'avons donc pas à remplacer les valeurs vides par une chaîne de texte.

#### 2.1.2 Diviser les cellules multivaluées

À partir de la colonne *Disponibilité*, allez dans `Editer les cellules > Diviser les cellules multivaluées`

Diviser les cellules multivaluées

Comment diviser des cellules à plusieurs valeurs

☒ par séparateur  
Séparateur  ☒ expression rationnelle 2.

☐ par les longueurs de champs

Liste les longueurs en les séparant par des virgules, par exemple 5, 7, 15


☐ par le passage des minuscules aux majuscules  
[11Abc, Def22]

☐ par le passage des chiffres aux lettres  
[11, AbcDef22]

☐ Inverser l'ordre de division  
[11A, bcD, ef22]

☐ Inverser l'ordre de division  
[11AbcDef, 22]

Indiquez comme séparateur \n (1.) et indiquez qu'il s'agit d'une expression rationnelle (2.). Validez.

 **A noter :** regarder le nombre de lignes (en haut à gauche de l'écran de projet), il est passé de 8 à 12. Passez maintenant en mode entrées. Notez que nous avons toujours 8 entrées.

### 2.1.3 Répéter les valeurs sur l'ensemble des lignes

À partir de la colonne *Toutes*, cliquez dans *Editer les colonnes > Recopier les valeurs dans les cellules vides consécutives*.

### 2.1.4 Extraction dxu nom de la bibliothèque

Regardez la construction du champ disponibilité (Vous pouvez pour cela vous aider des facettes). Le motif de construction du champ est toujours le même.

```
Disponibilité:Version physique à [BIBLIOTHEQUE] : [LOCALISATION]; [COTE HOLDING]; [COTE EXEMPLAIRE] ([DISPONIBILITE])
```

La bibliothèque est toujours précédée de la chaîne de caractère "Disponibilité:Version physique à " et suivie " : " qui introduit le reste des informations. Voici une des stratégies possibles pour extraire la bibliothèque :

- On supprime la chaîne "Disponibilité:Version physique à "
- On extrait ensuite la chaîne de caractères située devant " : " À partir de la colonne disponibilité, cliquez sur *Editer la colonne>Ajouter une colonne en fonction de cette colonne*.

Dans l'écran de saisie des expressions, saisissez la formule suivante :

```
value.replace("Disponibilité:Version physique à ","").split(" :")[0]
```

On remplace dans un premier temps la chaîne "Disponibilité:Version physique à " par une chaîne vide puis on utilise la fonction split qui va séparer notre chaîne sur la base de " :" en plusieurs chaînes de caractère. On prend alors la première chaîne ([0]).

Nommez la colonne *rcr* et validez.

2.3 Remplacer le nom de la bibliothèque par le RCR

À partir de la colonne *rcr*, cliquez sur *facette*>*facette* *textuelle*. Vous pouvez alors à partir de la



facette éditer chaque entrée pour la remplacer par le *rcr*.

Lib. Bib. Alma	RCR
Bib. STAPS	335222105
Bib. de l'institut du thermalisme - Dax	400882201
Bib. de l'IUT	331922101
BU SVS - Josy Reiffers	330632101

Supprimer la facette

3. Analyse de la réponse du web service

3.1 Voir si dans la réponse du web service le *rcr* de la colonne *rcr* existe

3.1.1 À partir de la colonne *réponse web service*, cliquez sur *Editer la colonne*>*Ajouter une colonne en fonction de cette colonne*.

3.1.2 Saisissez la formule suivante

```
filter(value.parseJson()["sudoc"]["query"]["result"]
["library"],bib,bib["rcr"]==cells["rcr"].value)[0]["shortname"]
```

Nommez votre colonne et validez.

Cette formule parcourt la liste des RCR. Si un RCR correspond au RCR présent dans la colonne RCR, elle retourne le nom de la bibliothèque dans le SUDOC.

3.1.3 Repérer les localisations non présentes dans le SUDOC.

À partir de la colonne nouvellement créée faites une facette pour repérer les cellules vides :

Facette>Facette personnalisée>Facette par valeur vide (valeur nulle ou chaîne vide) . La facette true vous permettra d'identifier les cas où la localisation dans le SUDOC est absente.