

# Chargeur Sudoc

Est reportée ci-dessous pour partie la documentation du chargeur Sudoc, que vous pouvez aussi retrouver sur le serveur dans le répertoire doc du chargeur, généralement ~/sudoc/sudoc/doc

## Import des notices : procédure de dédoublonnage

La recherche pour le dédoublonnage, c'est-à-dire la capacité de Koha à retrouver les bonnes notices locales à partir des notices sudoc, se fait à partir des index Zebra. Il est donc important pour que la moulinette fonctionne correctement, que les index Zebra soient correctement paramétrés. En particulier: que le numéro PPN ait été ajouté à l'index Identifier-standard.

Le dédoublonnage se fait comme suit:

- on teste d'abord la présence dans la notice Koha, en local, du PPN de la notice Sudoc entrante: le PPN sera toujours en 001 de la notice Sudoc, il peut être n'importe où dans la notice locale pour autant qu'il soit indexé dans l'index Identifier-standard (Ident).
  - Si la réponse est positive, on teste alors s'il y a lieu de signaler, dans le rapport de la moulinette, qu'il existe des notices fusionnées équivalentes côté Sudoc (signalées par l'abes par un 035\$aPPN, c'est-à-dire un 035 sans \$5 et avec un \$a qui n'est pas préfixé). La moulinette ne fait pas la fusion: elle la signale dans le rapport.
  - Si la réponse est négative, on teste alors la présence dans la notice Sudoc entrante du numéro de notice locale Koha. Entré en L035 dans le sudoc il est exporté dans la notice sudoc en 035, avec \$5 RCR et \$a numéro de notice local. On ajoute à ce test le même test sur les "doublons sudoc" que mentionné ci-dessus.

Les évaluations faites sont les suivantes:

- Evaluation 1 : le PPN de la notice sudoc correspond-il à un PPN connu localement dans l'index Ident?
- Evaluation 2 : existe-t-il dans la notice sudoc un 035\$a sans préfixe mais sans \$5? Permet de tester s'il existe un "doublon sudoc"
- Evaluation 3 : existe-t-il dans la notice sudoc un 035\$a sans préfixe avec un \$5? Permet de tester la présence d'un numéro de notice locale dans la notice sudoc
- Evaluation 4 : le 035\$5 trouvé précédemment correspond-il à un RCR du fichier rcr.yaml?
- Evaluation 5 : existe-t-il dans la notice sudoc un 930\$bRCR correspondant à un RCR du fichier rcr.yaml?
- Evaluation 6 : le 035\$a identifié correspond-il à un biblionumber local?

[Le graphe d'enchaînement des évaluations, et les différents cas qui en découlent \(mise-à-jour, création, rejet\) sont détaillés dans le fichier](#)

graphe\_dedoublonnage\_sudoc.pdf

## Détail des étapes

Les évaluations faites sont les suivantes:

### **Evaluation 1 : le PPN de la notice sudoc correspond-il à un PPN connu localement dans l'index Identifier-standard (ident) ?**

On recherche ici s'il existe dans Koha des notices dont le PPN (en 009) vaut le PPN de la notice sudoc (en 001), ou un ancien PPN sudoc s'il en existe (pris en 801\$h dans la notice sudoc)

#### **Si oui (éval 1) : une notice ayant ce PPN en 009 existe dans Koha**

Dans les logs on a, selon les cas :

Dans le fichier de log \*-a.csv :

```
000044105,"La notice existe dans koha avec les anciens ppn suivants:  
046377093",,,
```

```
000044105,"La notice a été trouvée par un ancien ppn: 046377093"," "," "," " "  
046377093,"1: maj",106875,"290192101 (BULET)","046377093 "
```

Dans le fichier \*.txt :

Identification sur le PPN

```
[08/10/2019 18:27:29] ----- Dédoublonnage de la notice 000003964
```

```
[08/10/2019 18:27:29] ident="000003964"
```

```
[08/10/2019 18:27:29] Cas 2 : mise à jour de la notice (ppn : 000003964 -  
biblionumber : 118264) - rcr : 290192103 (BUBOUG)
```

Identification sur un ancien PPN

```
[08/10/2019 18:28:17] ----- Dédoublonnage de la notice 000044105
```

```
[08/10/2019 18:28:17] ident="046377093"
```

```
[08/10/2019 18:28:17] La notice existe dans koha avec les anciens ppn  
suivants: 046377093
```

```
[08/10/2019 18:28:17] ident="000044105"
```

```
[08/10/2019 18:28:17] ident="046377093"
```

```
[08/10/2019 18:28:17] La notice a été trouvée par un ancien ppn: 046377093
```

S'il existe plusieurs notices dans Koha ayant le PPN recherché, aucune d'elles ne sera mise à jour, et dans les logs on voit :

```
[08/10/2019 18:29:52] ----- Dédoublonnage de la notice 000116955
```

```
[08/10/2019 18:29:52] ident="000116955"
```

```
[08/10/2019 18:29:52] Nombre de notices recues 2 différent du nombre de  
notices attendues (1)
```

[08/10/2019 18:29:52] Attention, doublons dans la base locale pour la notice de ppn 000116955 ! Je passe à la notice suivante

Si l'évaluation 1 est positive, c'est à dire qu'une seule notice a été trouvée dans Koha, alors on passe à la suite :

**Evaluation 2** : existe-t-il dans la notice sudoc un 035\$a sans préfixe mais sans \$5 = un numéro de PPN indiquant que la notice avec ce PPN a fusionné dans le Sudoc avec la notice entrante

**== Si oui (éval 2) Cas 1 : ==**

la notice Koha est mise à jour, et les logs indiquent les éventuels doublons

Recherche des doublons dans Koha :

Le test de dédoublonnage effectué à chaque mise à jour ou création de notice est le suivant :

- Si la notice Sudoc a un 035\$a sans préfixe et sans \$5
- Pour chaque 035\$a de ce type dans la notice Koha résultante (créée ou mise à jour), on recherche dans Koha les notices telles que ident = la valeur du PPN en 035\$a
- Si le nombre de ces notices est > 1, on affiche dans les logs le PPN retenu (celui de la notice Koha mise à jour ou créée) et les PPN rejetés (qui correspondront normalement à des doublons de notices)

Exemple avec doublon :

```
Dans le fichier de log *-a.csv on a :  
000055719,"La notice existe dans koha avec les anciens ppn suivants:  
011797878",,,  
000055719,"1: maj",73885,"290192101 (BULET)","011797878 "
```

Dans le fichier Sudoc entrant, on a la notice de PPN 000055719, qui possède plusieurs 035 sudoc :

```
035    _a011797851  
      _9sudoc  
035    _a012386960  
      _9sudoc  
035    _a008043434  
      _9sudoc  
035    _a098528351  
      _9sudoc  
035    _a061212458  
      _9sudoc  
035    _a072702036  
      _9sudoc  
035    _a011797878  
      _9sudoc  
035    _a021413762  
      _9sudoc
```

Dans Koha avant la mise à jour Sudoc, on a la notice de biblionumber=73885, dont le PPN est 000055719

Dans ce cas :

Evaluation 1 : recherche ident=000055719 (PPN sudoc) renvoie la notice Koha de même PPN, donc mise à jour de la notice.

Tests de dédoublement :

On recherche pour chaque 035\$a sudoc si une notice correspondante existe (ident=valeur du 035a)

Dans cet exemple, on a une réponse pour ident=011797878

En effet la notice koha de biblionumber=59561 (de PPN 011797878) est un doublon de la notice la notice de biblionumber=73885, dont le PPN est 000055719

Dans le fichier de logs détaillés ILN2019-10-08-01.txt on voit :

```
[08/10/2019 18:28:33] ----- Dédoublement de la notice 000055719
[08/10/2019 18:28:33] ident="011797851"
[08/10/2019 18:28:33] ident="012386960"
[08/10/2019 18:28:33] ident="008043434"
[08/10/2019 18:28:33] ident="098528351"
[08/10/2019 18:28:33] ident="061212458"
[08/10/2019 18:28:33] ident="072702036"
[08/10/2019 18:28:33] ident="011797878"
[08/10/2019 18:28:33] ident="021413762"
[08/10/2019 18:28:33] La notice existe dans koha avec les anciens ppn
suivants: 011797878
[08/10/2019 18:28:33] ident="000055719"
[08/10/2019 18:28:33] Cas 1 : mise à jour de la notice (ppn : 000055719 -
biblionumber : 73885) - rcr : 290192101 (BULET)
[08/10/2019 18:28:33] Regroupement des 955
[08/10/2019 18:28:33] Création des états de collection
[08/10/2019 18:28:33] Cette notice est un périodique ? : 0
[08/10/2019 18:28:33] Creating 099$s with 0
[08/10/2019 18:28:33] Creating 098$e with 2019-10-08
[08/10/2019 18:28:33] Updating 099$e with 2019-10-08
[08/10/2019 18:28:33] Début de rapport de dédoublement
[08/10/2019 18:28:33] Doublon sudoc : ppn retenu 000055719 - ppn rejeté :
011797878
[08/10/2019 18:28:33] Fin du rapport de dédoublement
[08/10/2019 18:28:33]
```

====> ceci permet d'identifier les doublons : dans le cas présent, la notice Koha de PPN 011797878 est un doublon

**== Si non (éval 2) Cas 2 : ==**

la notice Koha est mise à jour

**Si non (éval 1) : aucune notice ayant ce PPN n'existe dans Koha**

**Evaluation 3** : existe-t-il dans la notice sudoc un 035\$a sans préfixe avec un \$5? Permet de tester la présence d'un numéro de notice locale dans la notice sudoc (cas où le biblionumber Koha est saisi

dans la notice Sudoc)

### == Si oui (éval 3) : ==

la notice sudoc possède un 035\$a sans préfixe avec un \$5

**Evaluation 4** : le 035\$5 trouvé précédemment correspond-il à un RCR du fichier de configuration rcr.yaml (qui contient les correspondances RCR/branchcode)

### Si oui (éval 4)

**Evaluation 6** : le 035\$a identifié correspond-il à un biblionumber local

- Si oui (éval 6) Cas 3 : la notice Koha est mise à jour (ce qui est pertinent si le 035\$a est bien un numéro de notice Koha ; il arrive que ce soit les numéros d'anciens SIGB qui correspondent fortuitement à un biblionumber Koha, occasionnant la mise à jour d'une notice qui n'est pas la bonne) + Test pour voir s'il y a des doublons sudoc (Evaluation 2)
- Si non (éval 6) Cas 4 : la notice est créée dans Koha (ce qui pourrait être discutable, puisque le 035\$a ne correspond pas à un biblionumber ; mais ça peut être le numéro d'un ancien SIGB, donc dans ce cas ça a du sens de faire la création dans Koha) + Test pour voir s'il y a des doublons sudoc (Evaluation 2)

### Si non (éval 4)

**Evaluation 5** : existe-t-il dans la notice sudoc un 930\$bRCR correspondant à un RCR du fichier rcr.yaml : existe-t-il un exemplaire appartenant à l'un des RCR définis dans la configuration

- Si oui (éval 5) Cas 5 : la notice est créée dans Koha + Test pour voir s'il y a des doublons/fusions sudoc (Evaluation 2)
- Si non (éval 5) Cas 6 : la notice est écartée

### == Si non (éval 3) ==

: la notice sudoc n'a pas de 035\$a sans préfixe avec un \$5

**Evaluation 5** : existe-t-il dans la notice sudoc un 930\$bRCR correspondant à un RCR du fichier rcr.yaml : existe-t-il un exemplaire appartenant à l'un des RCR définis dans la configuration

- Si oui (éval 5) Cas 7 : la notice est créée dans Koha + Test pour voir s'il y a des doublons/fusions sudoc (Evaluation 2)
- Si non (éval 5) Cas 8 : la notice est écartée

## Commandes utiles

Pour extraire des logs des informations utiles, comme les doublons par exemple

```
Notices possédant un ancien PPN : identification des notices doublon dans Koha (non mises à jour) ou des notices trouvées par un ancien PPN (et mises à jour) :
```

```
cat ILN2019-10-08-01-a.csv | grep 'La notice existe dans koha avec les  
anciens ppn' > notices_avec_anciens_ppn.txt
```

Notices possédant un ancien PPN : uniquement les notices trouvées par un ancien PPN et mises à jour :

```
cat ILN2019-10-08-01-a.csv | grep 'La notice a été trouvée par un ancien  
ppn' > notices_trouvees_par_un_ancien_ppn.txt
```

Doublons de PPN dans Koha (donc notices n'ayant pas pu être mises à jour) :

```
cat ILN2019-10-08-01.txt | grep 'Attention, doublons dans la base locale' >  
doublons_dans_la_base_locale.txt
```

From:

<https://doc.bibliibre.com/> - **doc.bibliibre.com**

Permanent link:

[https://doc.bibliibre.com/koha/autour\\_de\\_koha/sudoc](https://doc.bibliibre.com/koha/autour_de_koha/sudoc)

Last update: **2019/12/27 17:47**

