

Imports des données SUDOC

Cas particulier : EEBO/ECCO

Le matching ne peut pas se faire sur ISBN ou ISSN. Il faut passer par les identifiants internes des plateformes. Ces identifiants dans le SUDOC se trouvent en 035 avec différents préfixes et dans Alma dans les paramètre d'analyseur des portefeuilles.

1. Extraction depuis Alma de la liste des Originated system id associé aux jkey (rapports /shared/Bordeaux NZ 33PUBD_NETWORK/docelec/listePortefoliosEEBO et /shared/Bordeaux NZ 33PUBD_NETWORK/docelec/listePortefoliosECCO)
2. Extraction depuis Alma de la liste des MMS id network (rapport /shared/Bordeaux NZ 33PUBD_NETWORK/docelec/listeMmsId à extraire au moins 2 fois pour obtenir la liste complète (les 400 000 premières lignes puis les 400 000 dernières lignes))
3. Extraction de winIBW de la liste des notices EEBO (che sou eeboIn? puis tel s1 k:003:035) ou ECCO (tdo o ; che sou eccoIn? puis tel s1 k:003:035)
4. Groupement dans OpenRefine des 3 fichiers pour obtenir une liste de la forme PPN - MMS_ID - Origin_id à extraire avec les noms de colonne suivants : 001-035\$a-24501\$a
5. Import dans Alma avec chargeur dédié (Import csv Marc21 - injection 035 - match 001 : cet import ne fait qu'ajouter le PPN en 035 sur la base du 001)
6. Extraction en MARC21 de la liste des PPN par MarcEdit via le batch
7. Injection des notices en marc21 (Import SUDOC MARC21)

A l'issu de ces opérations, de nombreuses imprécisions demeurent, notamment quand Alma associe plusieurs portefeuilles au même MMS_ID. La liste des anomalies doit être déposée ici après analyse.

From:
<https://reub-sgbm.frama.wiki/> - **Le wiki du SGBm du Reub**

Permanent link:
https://reub-sgbm.frama.wiki/new:administration_configuration:alma:docelec:importsudoc?rev=1556181468

Last update: **2019/04/25 10:37**

