

Horizon : migration des données bibliographiques

Les données bibliographiques sont migrées à partir du SUDOC. L'ABES réalise un export par établissement de toutes les notices sous lesquels les RCR de l'établissement sont localisés. Chacun de ces exports doit alimenter chaque zone institutionnelle (UB, UBM, INP, IEP). Pour l'alimentation de la zone réseau, l'agence nous fournit en plus un export de toutes les notices sous lesquelles les RCR de l'ILN 15 sont localisés. Un traitement permet

- d'ajouter de l'identifiant Horizon en 998
- de préfixer le PPN en 001 par « PPN »
- de s'assurer que la base bibliographique locale est correctement synchronisée avec le SUDOC.
- d'isoler les notices décrivant des documents électroniques pour lesquels il est nécessaire de réaliser une entrée dans l'inventaire électronique (P2E).

Périmètre des données exportées

Pour l'ensemble des notices

Toutes les notices fournies par l'Abes sous lesquels au moins un RCR du périmètre concerné est localisé. A l'exclusion des notices :

- Qui sont inconnues d'Horizon
- Qui sont présentes dans Horizon mais pour lesquelles aucun exemplaire n'existe dans Horizon pour les bibliothèques du périmètre exporté.

Ces notices exclues sont identifiées dans des rapports d'anomalies

Pour les notices signalées au P2E

Pour l'UB

Toute notice dont la localisation (930 \$5) dans le SUDOC correspond à un RCR électronique.

Pour UBM

Toute notice dont la localisation (930 \$5) dans le SUDOC correspond à un RCR électronique qui signalent des travaux universitaires (position 4 à 7 de la zone 105 \$a contient "m").

Pour le reste des établissements

Aucune notice candidate identifiée.

Méthode d'exportation

Script perl « **Migration_biblio_sudoc.pl** » sur le serveur de développement du SCOOP disponible sous “/home/scoopadmin/Workspace/Migrations/Migration_Horizon/Biblio/Migration_biblio_sudoc.pl” Le programme prend en entrée les notices fournies par l'ABES à déposer sous :
“/home/scoopadmin/IN_PERL/Preparation_Migration/Horizon/”

Analyse des anomalies

Lors de l'enrichissement des notices SUDOC le programme s'assure que base locale et base SUDOC sont correctement synchronisées.

Zone de l'analyse	Type d'anomalies détectées	Nom du fichier signalant les anomalies	Notices migrées
Au sein des notices fournies par l'Abes	Notices absentes de Horizon : le PPN n'existe pas dans Horizon	PPN_INCONNUS_HZ.csv	Notices non migrées
Au sein des notices fournies par l'Abes	Notices présentes dans Horizon pour lesquelles aucun exemplaire ou abonnement n'existe dans la base locale pour les bibliothèques du périmètre exporté.	LOC_ABSENTES_HZ.csv	Notices non migrées
Au sein des notices fournies par l'Abes	Notices pour lesquelles il existe plusieurs notices avec le même PPN dans Horizon	Doublon_ou_PPN_mal_saisi.csv	Notices migrées

Zone de l'analyse	Type d'anomalies détectées	Nom du fichier signalant les anomalies	Notices migrées
Au sein des de la base bibliographique Horizon	Notices ayant un PPN des exemplaires, des abonnements ou une localisation électronique dans Horizon mais dont la notice n'a pas été fournie par l'ABES	[Code de la division]_LOC_ABSENTES_SUDOC.csv	Notices et exemplaires non migrés

Les différentes listes des anomalies constatées ont été transmises aux responsables de catalogage. Elles peuvent être téléchargées [à cette adresse](#).

Le résultat de la migration de test

Vous trouverez dans ce tableau une synthèse des anomalies qui nous ont été signalées et les réponses qui y seront apportées.

Anomalie	Explication	Correctif à apporter
Absence des champs bibliographiques 109 (UB - PPN183587545)	Le champs 109 est un champs Horizon uniquement - comme les données proviennent du SUDOC, ce champs n'existe pas dans Alma	Aucun : c'est normal
Problème d'affichage des signes dans les mentions d'auteurs parallèles dans certaines notices (UBM - 193141183)	Contrairement au SUDOC, Alma n'attend pas dans la seconde valeur du \$f en 200 un signe de ponctuation car il en mettra automatiquement un, on a donc un affichage de 2 signes de ponctuations : un issu d'Alma et l'autre du SUDOC	Ticket en cours pour trouver une solution avec Ex Libris
Problème d'affichage des signes de ponctuation dans les titres parallèles (200 \$d commençant par) (UBM - PPN 056532202)	Comme ci-dessus : Alma met directement un signe '=', celui envoyé par le SUDOC est donc redondant	Ticket en cours pour trouver une solution avec Ex Libris
Les zones 120 et 121 n'ont pas toutes les sous-zones du SUDOC	Dans le SUDOC, ces zones comportent des sous-zones alors qu'en UNIMARC standard, ce sont des champs à position ⇒ les zones nous sont donc livrées par le SUDOC dans le format d'échange sous forme de champs à position, d'où leur import brut dans Alma.	Aucun, c'est normal.
Les zones 710 et 712 ont été inversées dans la notice (INP - PPN036094323)	La notice nous a été fournie telle qu'elle est dans Alma par le SUDOC : en réalité, la notice a été modifiée dans le SUDOC entre la date de notre extraction et la date de vérification ⇒ dans la base figée Horizon, la notice est identique.	Aucun, tout va bien

Anomalie	Explication	Correctif à apporter
Il manque le \$b dans la zone 105 (UB - PPN201428407)	Dans le format d'export du SUDOC, qui respecte le standard Unimarc, la zone 105 \$a est un champ à position. L'équivalent du \$b correspond aux positions 4-7. On les retrouve dans Alma dans les positions adaptées.	Aucun, tout va bien
Il y a des données surnuméraires en 130\$a (même exemple que ci-dessus)	Dans le format d'export du SUDOC, qui respecte le standard Unimarc, la zone 130 \$a est un champ à position. On retrouve donc dans Alma l'intégralité des sous-zone prévues pour la zone 130 dans la \$a telles qu'elles ont été transmises par le SUDOC.	Aucun, tout va bien
Les caractères de tri du SUDOC en zone 200 n'ont pas été repris	Il y a eu une erreur au chargement initial sur les données de tri. Ces données ont depuis été reprises et seront effectivement prises en compte à la migration finale : les caractères de tri sont modifiées et ne sont pas affichées dans l'interface.	Ex Libris : fournir le traitement à appliquer aux notices - fait
Pour une notice de partie composante, le type de matériel n'est pas bon (livre) (UB - PPN127250026)	Le type de matériel est déterminé par les données codées de la notice bibliographique. Un ajustement des données du LDR notamment est sans doute nécessaire. (détail du mapping)	Aucun
Absence de la zone 209 du SUDOC dans la notice Alma (UB - PPN127250026)	Cette zone, non-standard et spécifique à la gestion du PEB dans SUPEB n'est pas transmise par l'ABES dans ses exports.	Aucun
Absence des zones de lien (ex 440) dans les notices (UB - PPN036321761)	Les liens avaient été migrés en 440 et dans les autres zones mais sans activer les liens dans le système. 2 traitements sont nécessaires pour activer ces liens : un premier lié au traitement des données SUDOC qui avait été oublié par Ex Libris (rétabli depuis, c'est le même que celui qui gèrent les caractères de tri) et un second pour construire les liens. Ce dernier ne pouvait pas être activé immédiatement, il devait attendre les vérifications que nous avons à faire pour ne pas modifier les données livrées. Il a été activé depuis et le lien en 440 est bien disponible dans l'onglet autres détails de la notice.	Aucun
Mélange des exemplaires de périodiques (UB - PPN03922573)	Voir la page Migration des périodiques pour plus d'explications	
Absence des titres alternatif en 517 (UB - PPN036321761)	Les zones 517 sont bien présentes dans la notice Unimarc mais ne sont pas affichées dans la liste des résultats. Ils sont néanmoins bien recherchables.	Aucun
La zone E316 est reprise en 316 (UB - PPN185262864)	Cela vient du format d'export du SUDOC qui envoie la donnée en 316. Pour la migration initiale, nous aurons le même résultat. Nous travaillons par ailleurs sur le format d'envoi des exemplaires au SUDOC pour stocker cette donnée ailleurs. Plus d'explications sur les holding et les localisation SUDOC ici - attention, travail encore très en chantier !	Aucun pour le moment - chantier à mener post-migration pour la synchronisation des exemplaires.
La zone 210\$r n'est pas indexée dans Alma (UB - PPN098221949)	En effet, la zone 210 \$r ne figure dans aucun index de recherche professionnel.	Ticket ouvert auprès d'Ex Libris pour la gestion des index.

Anomalie	Explication	Correctif à apporter
Les notices liées ne sont pas cliquables (UB - PPN202394069)	Un traitement était nécessaire pour activer les liens entre notice : ils ne sont pas cliquables dans la notice mais tous disponibles dans l'onglet autres détails, notices liées.	Aucun

From:

<https://rebub-sgbm.frama.wiki/> - **Le wiki du SGBm du Rebus**

Permanent link:

https://rebub-sgbm.frama.wiki/projet:migration:horizon:hz_biblio

Last update: **2019/05/15 15:43**

