

Journée Systèmes et Données - 4 décembre 2020

Titre : **Notice bibliographique en arbre OEMI**

Tiphaine Cécile Foucher

Toutes les questions ont été posées via le chat lors de la journée. Certaines ont été posées en direct par des modératrices. Celles qui n'ont pu être traitées en direct ont été transmises aux intervenants pour une réponse par écrit.

Réponses apportées par écrit après la journée :

- Quand est-t-il prévu que le Sudoc propose les nouvelles zones à venir en 2021 ?
Non pas de calendrier pour le moment, nous y travaillons mais ce n'est pas prêt et nous n'avons pas les outils qu'il faut.

Réponses apportées au cours de l'intervention :

- Comment s'articulent les chantiers de la BnF "d'éclatement" des notices et ceux de "regroupement" (FRBRisation) pour créer des œuvres ?

Il s'agit de deux projets différents. La migration est liée au changement d'outil de production des données de la BnF (NOEMI) et RobotDonnées est un outil utilisé pour créer automatiquement des œuvres, les œuvres ainsi calculées sont visibles dans data.bnf.fr. RobotDonnées nous permet d'acquérir une expérience précieuse sur la création d'œuvre.

- Pourquoi la BnF ne profite pas de cette opportunité pour cataloguer en Unimarc ?
Quel est l'intérêt pour vous de conserver Intermarc ?

Quand la BnF a commencé à instruire le projet de changement d'outil de production des métadonnées, il a été décidé de créer un nouveau format de catalogage qui permettrait, dès l'ouverture de NOEMI, de cataloguer dans un format nativement compatible avec IFLA-LRM. De plus, les raisons historiques du choix initial de la BnF d'avoir son propre format de catalogage sont encore totalement d'actualité, à savoir : la complexité des types de documents et les obligations liées aux chaînes de traitement du dépôt légal qui imposent une réactivité rendant difficile l'adoption d'un format maintenu au niveau international.

- Quel format de référence la BnF va utiliser pour produire des entités ? Unimarc-LRM ?

La BnF produira des données dans son format interne de production (IntermarcNG). Elle continuera de diffuser ces données en Unimarc, qui est le format d'échange international défini comme format d'échange des bibliothèques en France. Le Comité français Unimarc travaille pour permettre à l'Unimarc d'exprimer un modèle de données compatible avec le modèle IFLA-LRM.

- Pourquoi faire un lien entre deux notices de monographie ayant le même titre ? Est-ce juste la différence d'édition, traduction, etc... Est-ce que ce sera lié à l'expression, du coup, plus tard ?

L'image présente dans le PPT est juste un exemple fictif. Il s'agissait simplement de montrer qu'à leur actuelle il y a déjà des liens riches entre les notices.

- Dans l'exemple 2, n'y a-t-il pas de relation entre l'œuvre agrégative et les œuvres qu'elle contient ?

Conformément à IFLA-LRM il n'y a pas de lien entre les œuvres agrégatives et les œuvres contenues. Quand l'agrégat éditorial est dépouillé il y a deux possibilités :

- la relation d'agrégation des expressions individuelles se fait au niveau de l'expression agrégative ;
- la manifestation agrégat est liée à toutes les expressions individuelles contenues (chaque expression individuelles est elle-même liées à son œuvre individuelle)

- Les processus de migration et autres se feront-ils de façon automatique (intelligence artificielle, algorithmes, etc) ?

A l'issue de la migration on estime que plus de 60 millions d'entités seront générées. Des traitements automatiques sont envisagés au même titre que l'utilisation de l'AI. Mais ce type de traitement ne réglera pas toute la migration qui est basées sur des règles métiers spécifiques pour chaque type de document.

- Combien de personnes travaillent à l'éclatement des notices ?

Le groupe est composé d'environ 18 personnes, les membres allouent à la migration entre à 10% à 40% de leur temps en fonction de leurs attributions dans le groupe. Le temps des membres du groupe pourra connaître des évolutions au cours du projet.

- N'y a-t-il pas de traitement de masse ?

Voir réponse à la question « Les processus de migration et autres se feront-ils de façon automatiques (intelligence artificielle, algorithmes, etc) ? »

- Est-ce que vous utilisez des outils particuliers pour la visualisation des liens ou le clusterisation des données?

A l'heure actuelle nous n'en sommes pas à cette étape-là, le premier cycle de migration n'est pas terminé donc nous n'avons pas encore débuté le travail sur la clusterisation des œuvres et des expressions.

- Qu'appelle-t-on clusterisation des données ?

Quand on parle de clusterisation on parle de fusion d'œuvre et d'expression. L'issue la migration on aura créé un arbre complet pour chaque notice bibliographique existante, il faudra ensuite rapprocher les entités qui décrivent des œuvres identiques et créer les clusters d'expressions.