

La récupération des données d'OGR : Excel2HAL

Préparer le fichier Excel

Grâce à l'aide de notre Direction des ressources informatiques, nous avons réussi à extraire les données de notre OGR dans un format lisible par Excel. Les informations bibliographiques étaient disposées en colonnes (titre, auteur, pagination etc.) et structurées selon les exigences de Bib2HAL (ex. : noms d'auteurs présentés sous la forme Nom1, Prénom1 AND Nom2, Prénom2).

Notre premier travail a consisté à rajouter des colonnes contenant les métadonnées absentes de l'OGR mais indispensables à toute intégration dans Bib2HAL, en fonction du type de document (ex. : audience, niveau de vulgarisation etc.). Quand certaines informations n'étaient pas connues (ex. : résumé et mots-clés), nous avons indiqué une valeur par défaut de type « Résumé à venir » ou « [Pas de mot-clé] ».

Nous avons ajouté une colonne « x-onbehalfof », dans laquelle nous avons entré manuellement, quand c'était possible, l'identifiant du compte HAL de l'auteur, retrouvé grâce à la requête mise au point par l'École des Ponts ParisTech. Il nous a en effet paru fondamental de partager la propriété d'un dépôt dès son intégration dans HAL, même si de nombreux chercheurs n'ont pas encore créé de compte sur la plateforme.

Nous avons également, après quelques tests, dû ajouter une colonne destinée à recevoir une « bibkey », soit un identifiant unique destiné à identifier la référence lors de l'intégration via Bib2HAL. Cette bibkey peut prendre n'importe quelle forme unique (ex. : bibkey1, bibkey2 etc.).

Générer un fichier XML

Avertissement : l'opération décrite ci-dessous a été mise au point par des bibliothécaires sans expertise dans le domaine, mais elle a le mérite de fonctionner...

Afin d'obtenir des informations bibliographiques structurées par des balises, nous avons entrepris de convertir nos fichiers Excel en fichiers XML, en adaptant [un tutoriel découvert sur internet](#). Cette manipulation consiste à créer un fichier de mapping XSD, en « appelant » chaque colonne par son en-tête. Chaque référence devant être précédée d'un couple de balises qui précise son type (article, book etc.), il est nécessaire de créer un fichier de mapping et un fichier Excel par type de référence à intégrer dans HAL (nos exemples sont téléchargeables dans le volet droit de cette page).

Dans un nouveau classeur Excel vierge, il faut ensuite utiliser les options de l'onglet Développeur pour ajouter le fichier de mapping XSD nouvellement créé en tant que source, et glisser-déposer l'architecture XSD dans la feuille de calcul. Ne reste plus ensuite qu'à copier-coller les données bibliographiques contenues dans votre premier classeur Excel, et à enregistrer le fichier ainsi obtenu au format .xml.

Convertir le fichier en BibTeX

Après avoir commencé par rechercher/ remplacer les balises XML par des balises bibTeX à la main à l'aide d'un éditeur de texte – procédure longue et fastidieuse car la moindre virgule a son importance -, nous avons sollicité un ingénieur informaticien de la BU pour mettre au point un script permettant la conversion automatique de ces balises. Par exemple, <title> devenait ainsi "title = {" et </title> devenait "},".

Ce script, baptisé XML2BIB et développé en Perl, est téléchargeable dans son format natif dans le volet droit de cette page, pour les utilisateurs de Linux.

Nous vous le proposons également au format exécutable pour Windows : il vous suffit de l'enregistrer sur votre ordinateur, de le lancer, et de vous laisser guider !

On obtient ainsi un fichier BibTeX entièrement compatible avec Bib2HAL. Seules manquent les étapes de dédoublonnage et d'affiliation des auteurs, à effectuer directement dans l'outil.

Mis à jour le 20 juin 2017

Télécharger nos modèles de schémas de mappage XSD

[Articles de revues](#)
[Chapitres d'ouvrages](#)
[Ouvrages](#)

[Directions d'ouvrages](#)
[Actes de colloques](#)

Télécharger XML2BIB

[Le script Perl](#)
[Le programme exécutable](#)

<https://openaccess.parisnanterre.fr/evenements-passes/de-bibtex-a-hal/alimenter-hal-grace-a-des-donnees-excel/>