

Compte-rendu de la réunion GT 3 - mardi 25 février 2016

Participants : **Nicolas Alarcon** (Université de La Réunion), **Frédérique Bordignon** (ENPC), **Alain Monteil** (INRIA), **Cécile Pierre** (Université Rennes 2)

1. État de lieux des outils d'import

Bib2Hal : L'outil est-il connu ?

2569 imports depuis la création de Bib2Hal, environ 10 697 items (dont 1508 en texte intégral, essentiellement INRIA)

Le support est assuré par l'INRIA (Laurence Farhi et Anne Jaigu)

L'outil est unanimement jugé très convivial et bien documenté

Import du texte intégral : possible si on indique l'URL, en déposant le fichier dans un endroit accessible (site web, FTP Hal lié au login, cloud sans authentification), mais pas directement pour les pdf dans zotero.

Idee d'adopter un fonctionnement plus proactif (cf dissem.in, Researchgate... qui proposent d'identifier des publications). Problème des affiliations pour Dissem.in + ORCID presque obligatoire.

Le dépôt du texte intégral demande d'indiquer toutes les affiliations, or les spécialistes IST connaissent bien les laboratoires de leur université mais ont des difficultés à identifier les labos extérieurs (risque d'erreur)

Lorsqu'on essaye d'affilier tous les auteurs, on se heurte tout de même au problème des doublons de structures (l'un des problèmes de l'import massif, notamment via prodinra, mais aussi en création)

Bib2Hal refuse de créer des structures non valides, effort préalable de création à faire le cas échéant dans le référentiel.

cela rejoint les préoccupations du groupe Référentiels (propositions, pont entre Aurehal et RNSR). Chaque responsable Hal doit faire le ménage pour son établissement et entrer le numéro RNSR pour chaque structure.

Problème de droit lorsqu'il n'y a pas de portail : il faudrait au moins qu'une personne puisse intervenir Aurehal par établissement.

Une formation à Bib2Hal devrait être organisée dans le cadre de l'UMS CCSD.

X2Hal : Suite de Bib2Hal. Import en TEI puis autres formats (Endnote, etc.)

Sword : utilisé par l'INRIA pour l'IFIP : métadonnées XML, feuille XSLT pour la conversion au format Hal. Nettoyage des référentiels a posteriori. Sword est utilisé aussi par

Polytechnique, (et BMC jusqu'à l'an dernier). Boucle, permet l'import d'une notice ou un import massif (pas de limitation de taille comme dans Bib2Hal)

Zotero n'est pas une solution adaptée pour les chercheurs : plutôt réfléchir sur l'expérience utilisateur lors du dépôt dans Hal

Gagner du temps avec des services souples (aspiration de références du Wos / scopus, possibilités d'export), éviter les multidépôts et multidéclarations.

Règles d'import

Bonnes pratiques à définir pour tous les services qui importent massivement, notamment concernant les structures

La clé RNSR est plus facile à gérer, mais il reste aussi du nettoyage à faire dans cette base (identification du correspondant de l'établissement)

Contact support RNSR : Xiaofeng CHEN + site web

https://www.galaxie.enseignementsup-recherche.gouv.fr/ensup/etab_RNSR.htm

L'inserm a réfléchi sur les structures, notamment pour les laboratoires à l'étranger et les CHU : à voir avec le groupe référentiels

Est-il bien ou non d'importer (souvent sans texte intégral) ?

Le passage d'une archive nationale à une archive d'établissement rend cela quasiment inévitable. Avoir un certain niveau d'exigence pour les données...

Nb : le texte intégral final du postprint n'est pas toujours accessible aux chercheurs, s'ils ne sont pas auteur référent ou s'il y a eu des échanges de dernière minute avec l'éditeur.

Du pragmatisme est nécessaire dans la gestion et dans le dépôt du texte intégral.

La modération est parfois très rigoureuse, avec des règles qui s'adaptent mal à certaines disciplines (dépôt du texte intégral avant conférence avec actes nécessaire en informatique, champ source obligatoire) ou strictes (ordre des auteurs) qui risquent de décourager les chercheurs

2. État des lieux des outils d'export

Pour l'instant l'outil d'export intégré à Hal n'est pas optimal, ce qui est dommage. Des outils extérieurs (ExtrHal) sont performant (format rtf, surlignage des auteurs du labo).

L'export Hal propose soit un export basique, soit un export avancé dans des formats qui n'intéressent pas les chercheurs (csv, xml, json).

HalTools : 2 Haltools, un format page web et un outil d'export (interne à l'INRIA actuellement) qui permet les exports bibtex, xml et endnote.

Les chercheurs demandent aussi un export en LaTeX

URL pour récupération automatique des différents types de documents pour rapports HCERES, idem pour l'API

Autres outils évoqués (non connus du groupe) : Aigle ; Graal puis Caplab (projets AMUE, probablement abandonnés).

Il serait souhaitable qu'un CRIS open source s'empare des données Hal.

Plugins d'interfaçage CMS/HAL:

- Wordpress : implémenté notamment dans les carnets Hypothèse
- Drupal (utilisé à Rennes)
- Typo3
- KSup (Toulouse 2)

Inria utilise l'outil décisionnel SpagoBI avec les données Hal pour des usages bibliométriques et d'indicateurs

AnHalytics : très expérimental, outil de textmining

Ces outils périphériques sont peu connus et assez difficiles à mettre en oeuvre

Propositions

- Intégrer un outil d'export clé en main à Hal, qui permette un export en RTF et/ou pdf (comme dans l'ancienne version de Hal)
- Rendre accessible à tous la version INRIA du module d'export Haltools
- Proposer des exports aux formats demandés pour les rapports fournis de façon récurrente par les laboratoires (HCERES notamment)
- Organiser des formations à l'API Hal pour les professionnels de l'information, qui ont besoin de ces compétences, par exemple sous forme de cours en ligne (un des enjeux du GT admin est de mettre en commun ces compétences)

Participation au GT 3 des informaticiens Onera etc souhaitée

Connexion avec GT 2, 4 et 5

Prochaine réunion : Jeudi 31 mars 2016 9h-11h