

Réunion GT 2 - jeudi 28 janvier 2016

Ordre du jour

1. Quel est le périmètre du groupe référentiels ?

Proposition : le périmètre du groupe est le suivant :

- référentiels d'AureHAL,
- interactions des référentiels d'Aurehal avec HAL,
- connexions/interopérabilité des référentiels d'AureHAL avec d'autres référentiels.

Le GT n'a pas vocation à traiter toutes les questions liées aux structures, aux formes auteur, etc. (au moins dans un premier temps).

ex :

- implémentation de dissem (GT4?)
- workflow de dépôt (GT1? ou GT 5?): *peut-on demander au ccsc un premier dépôt "minimum" (sans toutes les affiliations par ex) et compléter par l'admin portail ou référent labo ? A voir avec le GT5 ?*

2. Demander au CCSD des précisions sur le fonctionnement des référentiels et les interactions entre AuréHAL et HAL pour que l'on puisse mieux comprendre ce qui relève du fonctionnement normal et d'un éventuel dysfonctionnement

- ex : délais de mise à jour des formes auteur dans HAL après corrections dans AuréHAL*
- avoir des éléments précis sur les règles d'affichage des publications par auteur (pb décalage nbre publis associées à des formes auteur) -*
- indexation (non remontée de publis dans liste auteurs)*
- règles d'auto-complétion différentes pour saisie manuelle et récupération des métadonnées par le DOI, PMID, etc.*
- Impact des imports bib2hal et Star sur le référentiel "auteurs"*
- Niveau de droits/autorisations référents structure et admin portail >> à relier avec bonnes pratiques entre référents (changements à venir ?)*

3. Identifier les pbs et les dysfonctionnements et les hiérarchiser pour déterminer les plus urgents à résoudre à notre sens

- ex : imports massifs structures Prod'Inra (voir aussi la section "qualité des données des référentiels" ci-dessous)*
- ex : fonctionnement de l'auto-complétion pour les revues (signalé sur hal-infos)*
- ex : fonctionnement de l'auto-complétion pour les auteurs (affiliations associées : règles de proposition des affiliations)*
- ex : multiplication des formes auteurs concurrentes en dépit de l'IdHAL (ex. Ivica Closset signalé sur hal-infos)*

4. Identifier des demandes de développement et d'évolution et hiérarchiser leur priorité

- a. *ex : connexion avec les autres référentiels: Idref pour les auteurs et les structures, ORCID pour les auteurs, RNSR pour les structures, ISSN pour les revues, VIAF?, autres? Imaginer les points de convergence entre référentiels français (recommandations aux CCSD)*
 - i. *Question F. Gouzi : quid d'une future fusion avec l'IdRef (quand ?) auteur lorsque le premier sera stabilisé ?*
 - ii. *Question F. Gouzi : question de la stabilité et pérennité de l'idHAL (infos ou avancées à ce sujet côté ccsc ?)*
- b. *ex : développements d'outils/contrôles informatiques pour minimiser la création des doublons et faciliter le dédoublonnage :*
 - i. *détection/suggestion de doublons potentiels,*
 - ii. *pour les auteurs, afficher seulement les laboratoires des formes auteur à remplacer et non tous les laboratoires de tous les auteurs,*
 - iii. *pour les revues, recherche contient/titre exact/commence par (bon courage pour dédoubler "Nature" ou "Science")*
- c. *Prise en compte des structures nouvelles type COMUE (pertinent ou pas ?)*
Question F. Gouzi : possibilité de travailler au niveau local (COMUEs) et national, sur la mise en correspondance entre la notice ou id structure HAL (AurÉHAL) et l'identifiant IdRef « collectivité » correspondant, de façon à tendre vers une harmonisation des affiliations.
- d. *Historicisation des structures ? (date de début/de fin mais pb lien entre les structures). Lien avec le RNSR ?*

5. Comment mieux coordonner le travail des admins et des gestionnaires qui interviennent dans le référentiel ?

- a. *ex : pouvoir indiquer quelque part des règles de saisie concernant les unités de son établissement (code U et non UMRS pour les unités Inserm par exemple)?*
- b. *ex : demander au CCSD la possibilité de recevoir des notifications lors de la modification d'une forme valide de structure rattachée à son établissement ?*
- c. *Quelles règles quand on veut modifier une structure ? Prévenir le référent de la structure (du coup nécessite d'avoir la liste des référents à contacter ?)*
- d. *Définir règles de nommage des structures; Pb des structures homonymes nombreuses (cf Service de radiologie). Se mettre d'accord sur les champs à renseigner pour une structure (cf toujours mettre le n° RNSR) pour éviter les ambiguïtés.*

6. Quelles procédures imaginer pour améliorer la qualité des données des référentiels ?

- a. *problèmes de qualité des données identifiés :*
 - i. *identification de doublons de revues non encore validées, puis validation de ces revues ?*

- ii. détection et correction des noms de pays incorrects pour les institutions (le pays "France" semble avoir été mis par défaut pour de nombreuses institutions non validées)
 - iii. détection et correction des inversions de prénoms/noms d'auteur, qui proviennent d'un défaut d'ergonomie de l'interface de HAL lors du dépôt : les noms d'auteurs déjà présents sont indiqués dans l'ordre prénom puis nom alors que quand on entre un nouvel auteur, il faut commencer par le nom avant le prénom.
 - iv. identification des structures présentes plusieurs fois dans des langues différentes (par exemple un établissement allemand présent avec son nom en allemand, en français et en anglais)
- b. procédures possibles pour améliorer la qualité des données
- i. mise en place de normes pour certains aspects de certains référentiels (choix de la langue dans laquelle on indique le nom des structures validées, des critères qui justifient la création d'une nouvelle structure pour remplacer une structure obsolète : ajout d'une nouvelle tutelle, léger changement de nom, ou changement plus important de la structure ?)
 - ii. correction semi-automatisée de données du référentiel : possibilité de passer par des stages d'étudiants en informatique pour identifier les données erronées et proposer des corrections : mutualisation des problématiques et identification de laboratoires d'accueil possibles pour de tels stages ?

Pistes évoquées lors de la réunion d'octobre 2015

Pour A. Cartier, la gestion d'AureHAL est problématique. Pour le chercheur et pour les administrateurs, il y a un vrai travail de coordination à mener entre établissements.

Exemple des imports SWORD massifs effectués par Prod'inra : il n'y pas de contrôle à l'entrée des imports SWORD, ce qui crée des doublons. En fait Aurehal n'est pas un vrai référentiel : il n'y a pas de véritable interconnexion avec les données RNSR, ni avec les données ISSN pour les revues...

Il faudrait commencer par se mettre d'accord pour rentrer correctement les infos sur les structures.