

# Réunion GT 2 - jeudi 28 janvier 2016

## Compte rendu

Participants : Frédérique Flamerie (modératrice), Lucie Albaret (modératrice), Laurent Capelli, Anne Jaigu, Françoise Gouzi, Delphine Lescot, Isabelle Le Bescond, Claire Douady, Véronique Lacroix, Laurent Jonchère.

[Ordre du jour de la réunion](#)

## Informations et questions pratiques

L Capelli quitte ses fonctions de directeur technique au CCSD pour rejoindre l'équipe Isidore/Huma-Num (référencement des documents et données de la recherche SHS). Ses nouvelles activités restent cependant en lien avec les problématiques du GT2 (normalisation des référentiels), d'où sa participation au groupe de travail Référentiels.

- Les réunions du GT2 seront désormais programmées les mardis (voir calendrier).
- Outil de partage et de diffusion des documents du GT2 : Google docs
- Pour les réunions sur Skype, il est conseillé lorsque nous sommes nombreux de couper le micro afin d'éviter les bruits parasites.

## 1. Périmètre du groupe référentiels

Périmètre du GT2 restreint aux points suivants :

- référentiels d'AuréHAL,
- interactions des référentiels d'Auréhal avec HAL,
- connexions/interopérabilité des référentiels d'AuréHAL avec d'autres référentiels (ORCID, IdRef, etc).

Hors périmètre, mais en lien avec le GT2 :

- implémentation de dissem.in : lien avec les référentiels (GT4?)
- workflow de dépôt (GT1? ou GT5?)

## 2. Fonctionnement des référentiels et interactions entre AuréHAL et HAL

Il s'agit de demander au CCSD des précisions sur le fonctionnement des référentiels et sur les interactions entre AuréHAL et HAL afin que l'on puisse mieux comprendre ce qui relève du fonctionnement normal et d'un éventuel dysfonctionnement. Cela concerne particulièrement :

- le délai de mise à jour des modifications AuréHAL dans HAL,
- les différentes règles d'auto-complétion (auteurs, affiliations, revues), selon qu'il s'agit d'une saisie manuelle ou d'une récupération des données par un identifiant (PMID, arxiv...) ou par le DOI, ou encore par un import SWORD,

L Capelli : le CCSD serait en mesure de nous fournir une documentation sur tous ces points.

## 3. Identifier les problèmes / dysfonctionnements et les hiérarchiser

### Problème des correspondances structures

Les imports massifs de notices en provenance d'archives partenaires (Prondinra) génèrent dans AuréHAL des structures et des formes auteurs non valides. L Capelli précise que pour remédier à cet inconvénient, le CCSD a recommandé à l'INRA de ne plus affilier que les auteurs INRA, mais cette solution apparaît insuffisante.

De manière générale, de nouveaux développements seront nécessaires pour établir de façon automatisée des correspondances entre les référentiels HAL et les données auteurs / structures importées d'autres archives. Pour les structures françaises, la correspondance pourrait se faire sur la base du RNSR, choix d'autant plus pertinent que ce référentiel est alimenté et mis à jour par les directions de la recherche et intègre une hiérarchie (laboratoires / tutelles). Certaines universités ont déjà intégré les données RNSR dans AuréHAL. Ce système de correspondances est un des objectifs de la feuille de route du CCSD.

### Désambiguïsation des auteurs

Problème de la création de formes auteurs non valides. Typiquement, une forme auteur doublon est créée, par exemple via un import massif, alors qu'existe une forme ldHAL. Selon L Capelli, le nom et prénom ne sont pas suffisants pour attribuer automatiquement la

forme IdHAL, car il peut s'agir d'un homonyme. Il faut donc une donnée supplémentaire (adresse email, IdRef ou ORCID) pour désambiguïser.

La solution pourrait donc être d'associer systématiquement un ORCID et un IdHAL pour chaque auteur (il est déjà possible de renseigner l'ORCID en se créant un IdHAL). Mais si un auteur n'a pas créé d'IdHAL ni renseigné son ORCID ? l'administrateur portail pourrait « emprunter » le compte d'un auteur afin d'ajouter lui-même cette information, voire de créer l'IdHAL. L Capelli : pour le moment, cette possibilité est réservée aux administrateurs CCSD.

ORCID et idRef : ORCID est utilisé par les chercheurs au niveau international, alors que idRef est un référentiel français principalement utilisé par les bibliothécaires, d'où l'intérêt de privilégier ORCID comme pivot de désambiguïisation HAL.

Recommandations GT2 :

- demander au CCSD de pouvoir créer son idHAL à la création d'un compte,
- pour les nouveaux comptes HAL : récupérer l'ORCID,
- pour les comptes déjà créés : faire une recherche des ORCID et proposer aux auteurs de l'ajouter automatiquement, sur le même principe que le ResearcherID (quand un chercheur crée un compte ResearcherID, le système lui propose de créer un compte ORCID). Proposition en lien avec le GT4 Services aux chercheurs.

F Flamerie : HAL pourrait également envoyer des notifications aux auteurs, sur le modèle de ResearchGate, afin qu'ils puissent valider ou non telle forme auteur correspondante.

Décision GT2 : mise en ligne d'un document partagé pour recenser et hiérarchiser les problèmes et dysfonctionnements.

## 4. Coordonner le travail des référents AuréHAL

Etablir des règles communes

Constat : chacun travaille dans son coin. Il importe de se donner des règles communes de création de formes valides.

Exemple des CHU ou des laboratoires étrangers. Faut-il ou non ajouter la ville ? quelle langue privilégier pour un labo étranger ? L Capelli : HAL proposera bientôt une saisie multilingue des structures, ainsi qu'une historisation des structures (date de création, date de fin). Exemples de règles communes : ajouter le nom de la ville pour un CHU, ajouter l'identifiant RNSR, etc.

Décision GT2 : mise en ligne d'un document partagé pour recenser et hiérarchiser des règles communes.

Coordination des administrateurs ou référents

Constat : le travail de mise à jour des référentiels par les gestionnaires et administrateurs est perturbé par la création incessante de formes non valides, ou par la modification de formes valides correctes. Ce phénomène est accentué par les imports via webservices de notices en provenance d'archives partenaires (ex : Prodira). En conséquence, nombre d'administrateurs portail ont provisoirement cessé ce travail de mise à jour. Il faudrait pouvoir au minimum contacter par mail ou de manière automatisée le référent HAL d'une institution afin qu'il corrige ses structures.

Propositions GT2 :

- 2 référents maximum par institution, qui seront chargés de corriger les structures de leur institution (éviter les corrections multiples par différents acteurs). Etablir une liste CCD des référents institutionnels.
- CCSD : créer un service aux administrateurs d'aide à la fusion de formes valides : l'administrateur serait averti d'une modification et pourrait la valider d'un clic.

## 5. Identifier des demandes de développement et d'évolution et les hiérarchiser

Il faudrait simplifier au maximum la procédure de dépôt pour les chercheurs, afin de limiter la création de formes « parasites » dans les référentiels, et confier aux documentalistes et gestionnaires le traitement des métadonnées, en particulier des affiliations. La feuille de route 2016-2020 pour HAL (2016-2020) élaborée par le CCSD, qui sera prochainement diffusée, pourrait apporter des éléments de réponse sur ce point.

Cas des structures « atypiques ». Le système d'affiliation HAL ne permet pas de signaler correctement des structures de type COMUE et entreprises privées :

- COMUE : L Capelli évoque la possibilité de créer un référentiel COMUE au-dessus du référentiel des structures.
- Entreprises privées : selon L Capelli, le terme « institution » recouvre dans AuréHAL toute structure qui n'a pas de parents. Recommandation du GT2 : créer une catégorie « entreprise privée », ce qui permettra notamment de distinguer les partenariats privés dans une liste de publications.

Décision : mise en ligne d'un document partagé pour identifier les demandes de développement et d'évolution, et les hiérarchiser. Prévoir également une harmonisation avec les demandes des autres GT.

## Actions GT2 :

- Créer un document partagé : bonnes pratiques, règles communes
- Créer un document partagé (en lien avec autres GT) : problèmes et dysfonctionnements
- Créer un document partagé (en lien avec autres GT) : évolutions souhaitées
- Points spécifiques en lien avec d'autres GT :
  - GT Admin : le traitement des doublons de formes non valides (document support GT2 « bonnes pratiques »)
  - GT Services aux chercheurs : intégration des ORCID

Lien avec le GTA0

F Flamerie : les problématiques abordées dans le GT2 le sont également au sein du GT « Archives ouvertes » Couperin, nous ferons donc un retour des réflexions du GTA0 au sein du GT2, et inversement. La question des identifiants ORCID/idRef figure d'ailleurs à l'ordre du jour du prochain GTA0 programmé le 2 février.

## Calendrier des prochaines réunions GT2

Principe : organiser des réunions thématiques par référentiels.

Dates des prochaines réunions :

- 23 février : référentiel structures, point sur les contributions au document « [Règles de saisie des structures dans AuréHAL](#) »
- 29 mars : référentiel des revues
- 26 avril
- 24 mai
- 28 juin