

Compte-rendu de la Réunion OCA/SCD autour de HAL et du signalement des publications de l'OCA du 18 janvier 2019

(9h30 - Obs. de Nice - Pavillon Evry Schatzman - Salle Jean Arnaud)

Présents :

Ghislain Chave (SCD UNS), Catherine L'Hostis, Marie-Laure Miniussi, Catherine Renaud (Publications Lagrange),
Mathieu Saby (SCD UNS), Jenny Trevisan (publications Géoazur), Sylvie Vaquié
Excusé : Gilles Bogaert (publications Artémis)

Rappel de l'ordre du jour

1. Actualités de la science ouverte
2. Situation à l'OCA et besoins exprimés
3. Projet Conditor
4. Projet UCAGate
5. ORCID
6. Scopus et Scival
7. HAL et outils liés
8. Dissemis
9. Conclusion

En préambule : une première rencontre a eu lieu 3 ans auparavant (15 avril 2016) autour de cette thématique entre Ghislain Chave, Mathieu Saby et Géraldine Geoffroy pour le SCD et David Andrieux (Artemis), Gilles Bogaert (Artemis), Jelena Giannetti, Catherine L'Hostis, Bernard Mercier de Lepinay (Géoazur), Catherine Renaud (Lagrange), Pierre-Louis Sulem (Lagrange), Marie-Paule Torre, Jenny Trevisan (Géoazur), Sylvie Vaquié pour l'OCA

Nouveautés depuis cette réunion :

- **Le portail HAL UNS est devenu HAL UCA**, toujours administré par G. Chave
- **Le CCSD a publié une feuille de route pour l'évolution de HAL en sept. 2016** https://www.ccsd.cnrs.fr/wp-content/uploads/2016/06/Feuille_de_route_CCSD_2016-2020.pdf
- **Au niveau national, une association des utilisateurs de HAL a été créée en 2016** (<https://wiki.ccsd.cnrs.fr/wikis/hal/index.php/Accueil>). G. Chave y représente le SCD. Deux journées d'étude ont été organisées
- **L'UCA s'est abonné à Scopus et à SciVal**, essentiellement pour les besoins de la gouvernance, mais les membres y ont accès. Pour répondre aux besoins de l'Office of Scientific Visibility, le SCD a travaillé avec le fournisseur (Elsevier) pour améliorer l'identifications des publications UCA

1. Actualités de la science ouverte

Les activités du SCD s'inscrivent désormais dans le cadre du soutien à la « science ouverte » (open access, partage des données de la recherche...). Le contexte international, national et local évolue très rapidement depuis quelques mois. L'ANR impose de nouvelles exigences, qui devraient être précisées dans les mois à venir.

Dossier suivi par M. Saby, G. Chave et J. Pollet

2. Situation à l'OCA et besoins exprimés

Les laboratoires de l'OCA suivent chacun leurs publications avec des méthodes et des outils variés. Comment gagner en efficacité et en cohérence afin de répondre aux besoins, en particulier la production de rapports pour les différentes tutelles ?

L'open access n'est pas un objectif en soi.

L'existence d'une base de données interne recensant toutes les publications n'est pas un objectif en soi.

3. Situation

- Géoazur : J.Trevisan saisit les notices et parfois les documents dans HAL, et en parallèle renseigne une base interne pour répondre aux enquêtes demandées par les tutelles (ex : UNS demande des rapports prenant en compte le genre des chercheurs...) et pour alimenter la liste des publications sur le site du Géoazur. La base est un outil « maison » en MySQL.
- Lagrange : publications gérées dans Endnote. Exports vers la base du Géoazur. Jusqu'en 2012 un export vers HAL était réalisé, mais ce n'est plus le cas.
- Artemis : s'appuient principalement sur les dépôts faits dans HAL par les équipes de l'IN2P3 (articles avec souvent de très nombreux auteurs). G. Bogaert contrôle ces notices mais n'a pas d'autres outils. Pas d'exports vers la base du Géoazur.
- UMS Galilée : pas de suivi (?)

Pour veiller à la complétude du signalement, les collègues chargés du suivi utilisent d'autres outils (base d'astronomie ADS, WoS...)

Rappels de la réunion de 2016 :

« A l'OCA 2 archives ouvertes sont particulièrement utilisées : ArXiv et HAL. ... HAL ne fait pas l'unanimité parmi les chercheurs : trop compliqué, trop long. Le problème des affiliations est rappelé. S'il y a 1000 auteurs, il faut les rentrer un par un. Il ne permet pas non plus de faire de la bibliométrie. »

Point de vu confirmé aujourd'hui : les chercheurs ne voient pas l'intérêt de déposer dans HAL si leurs articles sont visibles et consultables ailleurs...

4. Projet Conditor

Projet national piloté par la DIST-CNRS, avec la participation de plusieurs établissements, suivi par J. Pollet au SCD. L'UNS ne fait pas partie du GT, mais d'un 2^e cercle d'établissements intéressés par les résultats du projet.

Conditor vise à produire un recensement consolidé des publications de la recherche française. L'échéance prévue était mars 2019, mais le projet se heurte à la difficulté juridique et économique de réutiliser les données du WoS et de Scopus, qui sont indispensables à son succès. Pour l'instant seules les données en accès libre (HAL, theses.fr, archives ouvertes locales) ont été prises en compte.

Conditor prévoit un dédoublonnage automatique des publications, et une affiliation semi-automatique de chaque auteur à une structure de recherche identifiée dans le Répertoire National des Structures de Recherche.

Si le projet aboutit, un travail de la part des établissements sera nécessaire pour valider des affiliations proposées par l'outil.

Conditor n'a pas prévu d'exploiter les identifiants ORCID.

Donc : à suivre mais ne semble pas répondre au besoin dans l'immédiat.

5. Projet UCAGate

Projet local UCA initié par la gouvernance, suivi par J. Pollet au SCD.

Semble définitivement stoppé, après la réalisation d'un prototype. L'objectif était d'analyser la production des chercheurs UCA pour en dégager des « compétences ». HAL aurait pu constituer une première base pour identifier ces publications, mais un recours à des sources complémentaires aurait été nécessaire.

Donc : à suivre si le projet est relancé, mais ne répond pas au besoin dans l'immédiat.

6. ORCID

L'UNS vient d'adhérer au consortium ORCID en 2019, avec une vingtaine d'autres universités françaises. J. Pollet est le contact technique.

Pour l'instant pas de projets concrets mais **l'adhésion nous donne accès à une API en écriture complémentaire de l'API en lecture (gratuite), et la possibilité d'interfacer ORCID avec d'autres applications.**

Question sur l'intérêt d'ORCID : ORCID permet d'enlever les ambiguïtés sur les auteurs homonymes et d'attribuer à chacun la paternité de ses travaux. Il permet de rattacher à un auteur des travaux qu'on ne pourrait pas trouver autrement (ouvrages publiés, données de la recherche dans Pangaea ou Zenodo, etc). C'est un système gratuit et dont la gouvernance est transparente. Il n'est pas la propriété du WoS, d'Elsevier ni d'un éditeur contrairement au ScopusID ou au ResearcherID. Il donne la main au chercheur pour ajouter ou supprimer toutes les informations qu'il souhaite et choisir leur degré de visibilité. Il ne remplace pas ScopusID, ResearcherID, IDHAL ou les identifiants créés par les bibliothèques mais permet de les lier de manière contrôlée.

Donc : démocratiser l'usage d'ORCID et d'IDHAL, en liant les deux, pourrait permettre la mise en place de services à valeur ajoutée, mais ça ne pourrait être qu'un projet à moyen terme, et complémentaire d'autres approches.

7. Scopus et Scival

L'UCA s'est abonné à la base bibliométrique Scopus et à l'outil d'analyse lié SciVal. L'objectif est de répondre aux besoins de la gouvernance (production d'indicateurs). L'accès est en théorie possible pour toutes les personnes affiliées à un établissement membre d'UCA. En pratique les accès hors UNS ont du mal à être déployés techniquement.

La production d'indicateur est du ressort de l'Office of scientific valorisation, service d'UCA en cours de reconfiguration. Mais le SCD (J. Pollet + E. Beauville) est référent technique.

Un travail important a été fait par le SCD pour reproduire dans Scopus la structure d'UCA : membres et laboratoires. Il est donc possible d'identifier les publications de l'OCA et de ses laboratoires. Les références peuvent être exportées dans différents formats ; un accès par API est également possible.

Donc : **Scopus peut être une bonne source pour suivre les publications de l'OCA, mais la base n'est pas exhaustive** (même si elle est plus large que le WoS), et ne comprendra pas toutes les revues, ni les travaux non publiés dans des revues (préprints, rapports, jeux de données, brevets etc).

8. HAL et outils liés

Depuis janvier 2018, le portail HAL UNS est devenu HAL Université Côte d'Azur <https://hal.univ-cotedazur.fr/>, toujours administré par Ghislain Chave. Début 2018, Marie-Laure Miniussi et Ghislain Chave ont travaillé ensemble sur la visibilité des publications de l'OCA dans HAL Université Côte d'Azur :

- Tous les dépôts de l'OCA sont tamponnés par la collection UNIV-COTEDAZUR et apparaissent dans le portail hal.univ-cotedazur.fr.
- Les laboratoires de l'OCA sont affichés dans la consultation par structure.
- Dans le référentiel AuréHAL, l'institution OCA est rattachée au regroupement d'institutions UCA. Les unités associées à l'OCA restent rattachées dans HAL à leurs différentes tutelles :
 - UNS, OCA, CNRS pour Artemis et Lagrange
 - UNS, OCA, CNRS et IRD pour Géoazur
 - OCA, CNRS pour Galilée

Le SCD propose un accompagnement mais ne peut actuellement consacrer qu'un temps limité à HAL (un peu plus d'1/3 ETP) et ne peut donc prendre en charge ou mettre en œuvre des services qui nécessiteraient davantage de ressources humaines, techniques ou financières.

Le CCSD a publié une feuille de route, et prévoit des évolutions dans les mois et années à venir pour faciliter l'usage de HAL. Depuis peu il est possible de déposer du code informatique, etc.
Le club utilisateur CASUHAL (<https://www.inria.fr/casuhal/>) fédère désormais les établissements utilisateurs. G. Chave suit ses travaux. Deux journées d'études ont été organisées, des outils et des méthodes y sont échangés, etc.

Une référence peut être importée dans HAL en quelques clics à partir d'un identifiant DOI ou d'un identifiant ArXIV. Les métadonnées peuvent également être extraites à partir d'un fichier déposé.

Des outils externes permettent de réaliser des imports semi-automatiques à partir de listes de références extraites de différentes sources. Cf la présentation <https://www.inria.fr/casuhal/files/2018/06/OverHAL-X2HAL-Dijon-2018.pdf> centrée sur deux outils :

- **overHAL** (développé par Rennes 1) <https://halur1.univ-rennes1.fr/OverHAL.php> : Permet de créer un fichier XML-TEI importable dans HAL à partir d'un fichier CSV exporté de Scopus, du WoS, de SciFinder, de PubMed ou de Zotero ; récupère les PDF si fichiers sous licence CC.
- **x2HAL** (développé par INRIA) <https://x2hal.inria.fr/> : Remplace l'application bib2hal. A partir d'un fichier XML-TEI créé par overHAL, calcule les métadonnées manquantes et les affiliations, repère les doublons puis crée les notices dans HAL et dépose les PDF.

Remarque : Un travail manuel peut être nécessaire pour corriger les affiliations et compléter les métadonnées. Des outils permettant d'exporter les données de HAL (API...) existent également.

Donc : les outils développés à INRIA, Rennes ou ailleurs pourraient mériter une exploration plus avancée, mais ils ne permettent pas une automatisation complète.

9. [Dissemin](#)

L'outil **Dissemin** (<https://dissem.in/>) bien que non conçu spécifiquement pour HAL, **permet de faciliter le dépôt de documents en texte intégral. Il facilite le repérage des publications juridiquement déposables dans une archive ouverte, et automatiser leur dépôt.**

Les archives ouvertes supportées sont HAL, Zenodo et OSF. **Dissemin cherche tous les articles disponibles dans CrossRef (articles dotés d'un DOI) et dans BASE** (moteur de recherche recensant des documents disponibles en open access issus de revues ou d'archives ouvertes), et trie les résultats en fonction du statut du document.

Dissemin doit être utilisé par un chercheur en son nom propre, qui doit s'identifier avec son ORCID. Il faudrait vérifier si le couplage ORCID/IDHAL est utile ou nécessaire à son bon fonctionnement.

Donc : *c'est un outil potentiellement intéressant pour améliorer le taux de documents en open access, mais il ne répond pas au besoin d'amélioration du référencement des publications de l'OCA.*

10. Conclusion

Réflexion suite à la réunion :

4 possibilités, la B me semblerait la plus réaliste même si c'est du boulot. Elle serait à coupler avec une stratégie d'enrichissement de l'Annuaire UCA et de généralisation des ORCID et des IDHAL. Dans ce cas il faudrait voir de plus près ce qui se fait à Rennes ou à Bordeaux, en terme techniques et de RH...

A/ Gestion des publications dans une base interne, alimentée par différentes sources (HAL, Scopus...)

- a. Base OCA
- b. Base UCA

Avantage : permet de produire directement des rapports, de coupler la base au reste du SI, etc.

Inconvénient : demande travail d'ingénierie documentaire et informatique, à financer ; besoin de RH pour l'alimentation et le contrôle

B/ Gestion des publications dans HAL, avec alimentation de HAL à partir d'autres sources (Scopus etc) pour l'OCA et/ou d'autres membres ou laboratoires.

Avantage : l'enrichissement de HAL en notices pourrait ensuite servir pour faciliter le dépôt des documents en texte intégral ; pas besoin de développer d'outils

Inconvénient : besoin de RH pour l'alimentation et le contrôle ; certains types de productions ne pourront pas être gérés avec HAL (données de la recherche essentiellement) ; les rapports devront être fait hors de HAL

C/ Pas de gestion des publications : attente d'une solution nationale (Conditor)

D/ Pas de gestion des publications : exports occasionnels à partir de Scopus, HAL, Arxiv, ADS, et analyse au cas par cas par chaque labo

Avantage : autonomie de chaque labo ; pas besoin de RH pour l'alimentation et le contrôle des bases ;

Inconvénient : bricolage, pas fiable, pas capitalisé ; besoin de RH pour le recoupement des bases et leur nettoyage avant l'analyse]

Compte-rendu rédigé par Mathieu Saby et Ghislain Chave