OverHAL-Zip2HAL

Atelier Journées CasuHAL 2025

Université de Rennes

Plan de l’atelier

1. **Contexte** : « cataloguer » les publications de son labo, de son établissement
2. **Pratique** : référencer un lot de publications OpenAlex dans HAL
3. **Retour en détail** sur les 3 étapes du référencement dans OverHAL-Zip2HAL
4. **Echange** : que vous inspire cet atelier ?
5. **Soulevons le capot** : comment travaille l’algorithme Zip2HAL ?

Posez vos questions à tout moment

Le « catalogage » des publications

Les références sont créées dans HAL

OpenAlex

PubMed

Web of Science

CSV

OverHAL

XML

Zip2HAL

HAL science ouverte

OverHAL : les publications (métadonnées) sont dédoublonnées, puis converties au format XML TEI-HAL

Zip2HAL : les métadonnées sont automatiquement enrichies via les référentiels HAL (ROR, ORCID…) puis déposées dans HAL via les webservices CCSD (Sword)

Source : poster présenté lors des Journées CasuHAL 2024 (extrait)

Atelier : importer un lot de notices dans HAL

5-6 notices maximum dans le zip (26 connexions simultanées !)

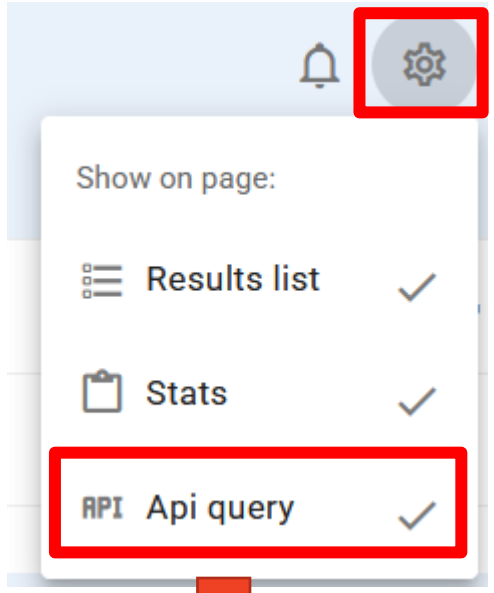
Etape 1 : générer une requête API dans **OpenAlex** (labo + année 2025)

Etape 2 : générer les XML HAL des notices dans **OverHAL** à partir des notices OpenAlex

Etape 3 : vérifier les notices XML dans **Zip2HAL** avant de les importer dans HAL

OpenAlex

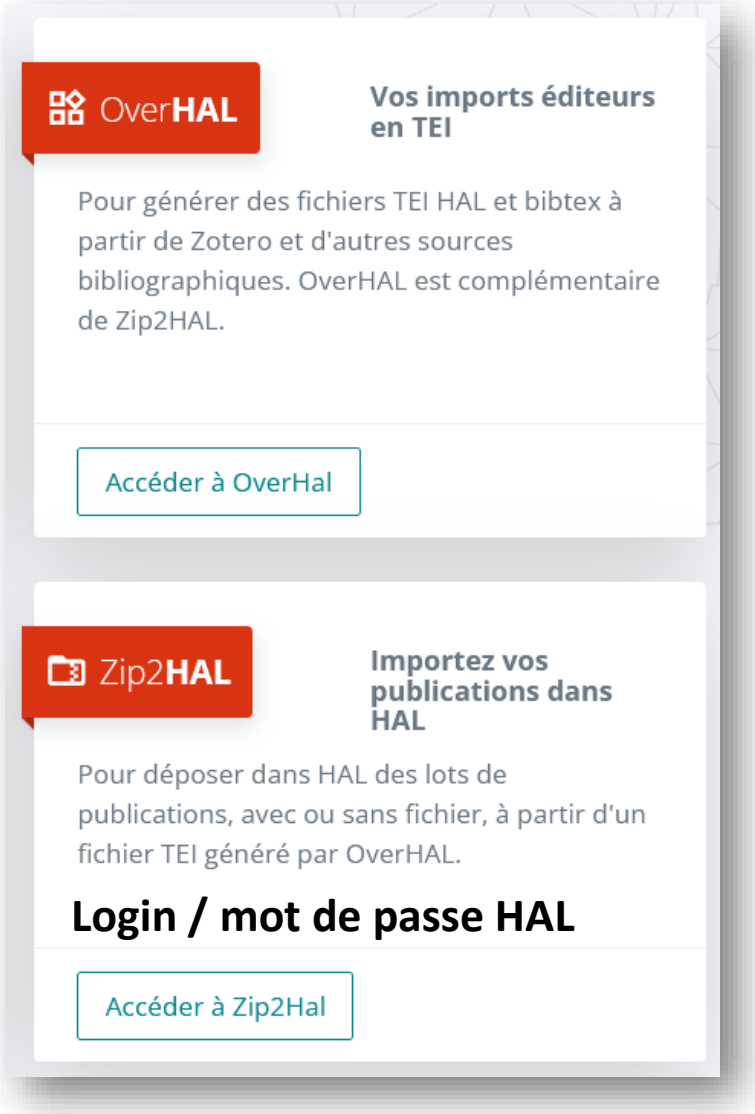
https://openalex.org/



API https://api.openalex.org/works?page=1&filter=authorships.institutions.lineage:i4210105651,publication_year:2025&sort=cited_by_count:desc&per_page=10
&apc_sum=false&cited_by_count_sum=false

HALUR

https://halur.ccsd.cnrs.fr/



Université de Rennes

Etape 1 : Générer les fichiers source

Construire ses requêtes OpenAlex

- Le plus simple : générer sa requête API dans OpenAlex

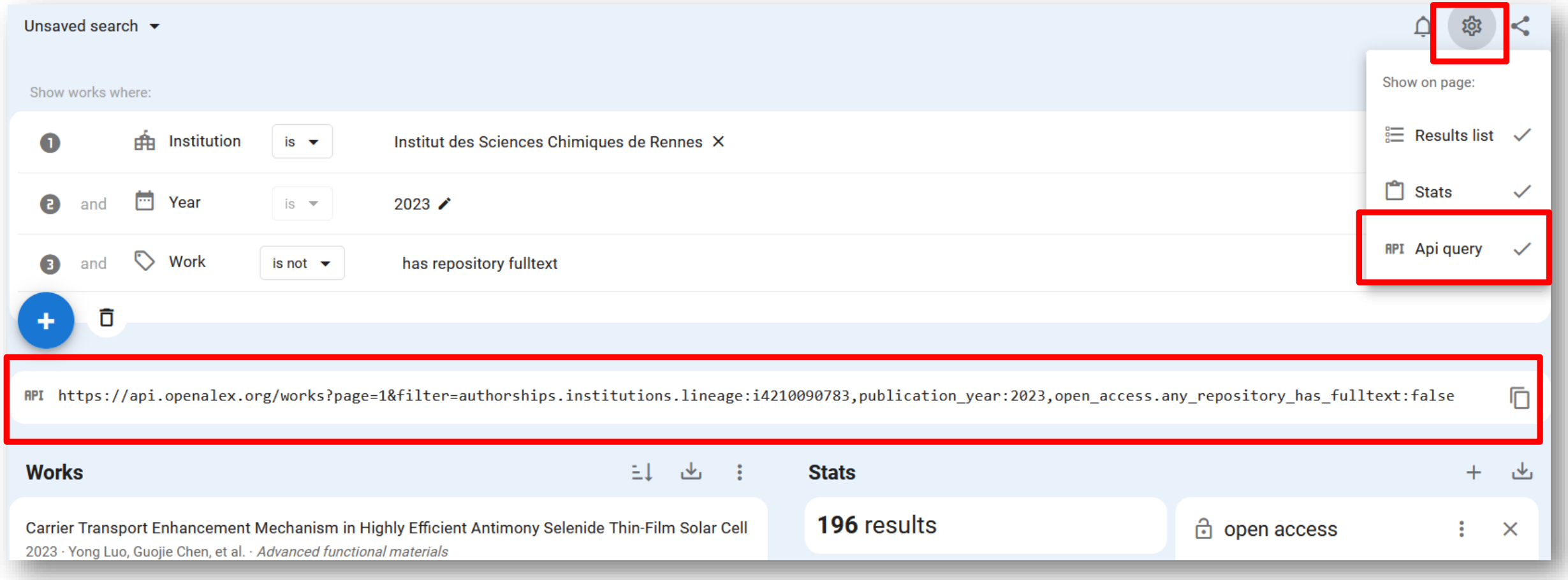
Pour une méthode plus complexe (requête sur les **affiliations brutes**) :

- **Prendre son temps**, étudier la documentation API : https://docs.openalex.org/
- Partir d'exemple concrets, à partir de DOI pour repérer les différents champs (affiliations, auteurs…)
  Exemple : https://api.openalex.org/works?filter=doi:10.1111/jce.16139
- Construire sa requête en s'aidant des diapos suivantes, pas à pas, et tester l'url dans un navigateur
- Faire l'effort de déchiffrer les éléments de la requête et des champs API pour devenir autonome dans la construction de ses propres requêtes
- Possibilité d'utiliser un éditeur comme Bruno pour construire ses requêtes

Construire la requête AI depuis l'interface OpenAlex

OpenAlex

https://openalex.org/



Inconvénient : il peut y avoir des erreurs d'affiliations (calculées à partir des formes brutes)

Exemple d'une notice d'article OpenAlex

https://api.openalex.org/works?filter=doi:10.1016/j.cris.2024.10007


```
author_position: "first",
author: {
    id: https://openalex.org/A5077419455,
    display_name: "Joan van Baaren",
    orcid: https://orcid.org/0000-0002-8552-9645
},
institutions: [
    {
      id: https://openalex.org/I4210087209,
      display_name: "Ecosystèmes, Biodiversité, Evolution",
      ror: https://ror.org/005fjj927,
      country_code: "FR",
      type: "facility",
```


L'algorithme qui calcule l'affiliation (display_name) n'est pas toujours fiable


```
is_corresponding: true,
raw_author_name: "Joan van Baaren",
raw_affiliation_strings: [
    "Université de Rennes, CNRS, ECOBIO (écosystèmes, biodiversité, évolution) - UMR 6553, 263 Avenue du Général Leclerc,
    35042 Rennes, France"
]
```

Exemple d'une notice d'article OpenAlex

https://api.openalex.org/works/doi:10.1016/j.saa.2022.121833


```
author_position: "middle",
author: {
    id: https://openalex.org/A5045325267,
    display_name: "Xiaoyan Yang",
    orcid: https://orcid.org/0000-0001-8508-4300
```

```
],
is_corresponding: false,
raw_author_name: "Xueying Yang",
raw_affiliation_strings: [
    "Key Laboratory of Rare Earth Optoelectronic Materials
    and Devices, China Jiliang University, Hangzhou 310018,
```


L'algorithme qui calcule le nom de l'auteur (display_name) n'est pas toujours fiable

Et donc l'ORCID n'est pas le bon…

Construire les requêtes API OpenAlex

https://api.openalex.org/works?**filter=raw_affiliation_strings.search**:UMR6553 OR "UMR 6553" OR "Écosystèmes biodiversité évolution" OR (ECOBIO AND France) OR "Ecosystem Biodiversity Evolution",**type:**article|book-chapter|book|editorial|letter|other|review|standard|retraction,**type_crossref:!**posted-content,best_oa_location.source.id:!https://openalex.org/S4306402512,**primary_location.source.type:!**repository,**from_publication_date:**2023-01-01,to_publication_date:2025-12-31,***from_created_date:****2024-01-01*&***api_key***=XXXXXXXXXXXXXXXXXXXXXXX&**mailto**=toto.titi@univ-rennes.fr

filter=raw_affiliation_strings.search: On recherche des éléments d'affiliation dans une chaîne de caractères « brute »

type: On limite notre recherche à certains types de documents (« articles » englobe les communications) https://api.openalex.org/works?group_by=type

type_crossref:! On exclut (!) de notre requête toutes les publications déjà référencées dans HAL (ou dans une archive ouverte **primary_location.source.type**)

from_created_date: On filtre sur les entrées récemment créées (uniquement avec une clé API **api_key** optionnelle). Attention, la date de publication dans OpenAlex n'est pas toujours fiable > faire une requête élargie AAAA

On peut insérer d'autres filtres dans la requête, exemple un auteur : display_name: Jean Dupont

En vert : les éléments à mettre à jour

https://docs.openalex.org/

Requête booléenne (affiliation)

OpenAlex

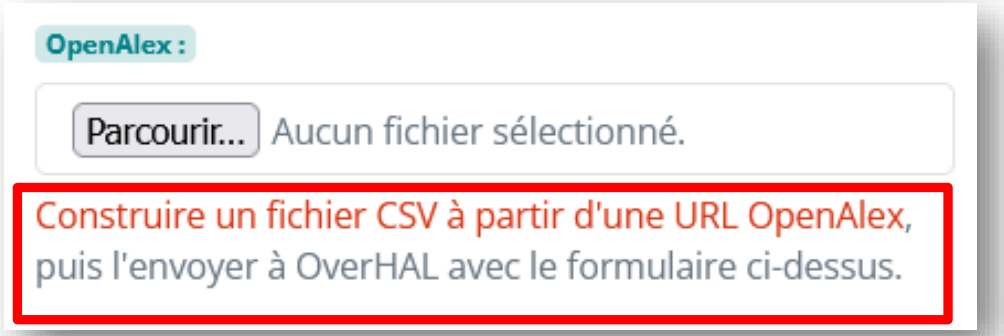
filter=raw_affiliation_strings.search:

(UMR6251 **OR** "UMR 6251" **OR** IPR **OR** "Institut de physique de Rennes" **OR** "Institute of Physics of Rennes" **OR** "Rennes Institute of Physics" **OR** LARMAUR **OR** "Mécanique et Verres"**) AND (**France**) NOT (**"Institut Pierre Richet" **OR** "Intelligent Process Automation and Robotics" **OR** "Sant Joan de Déu" **OR** "Institut Polytechnique Rural de Formation et de Recherche Appliquée"**)**

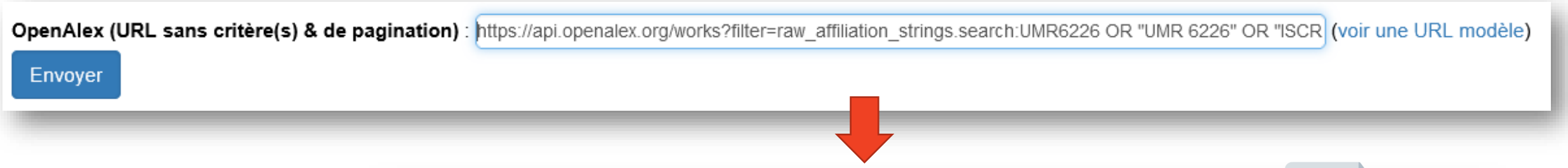
- Recherches booléennes : https://docs.openalex.org/how-to-use-the-api/get-lists-of-entities/search-entities#boolean-searches
- Ne pas mettre de virgule dans les chaînes entre guillemets (ou alors la rejeter) : "Oncogenesis**,** Stress and Signaling"
- On ne peut pas rechercher (" Institut de Physique" AND Rennes) entre parenthèses - mais ça fonctionne si on ne met pas de guillemets : (Institut de Physique AND Rennes)
- Uniquement des guillemets doubles
- Les troncatures et opérateurs de proximité ne fonctionnent pas (* ? ~ NEAR ADJ)

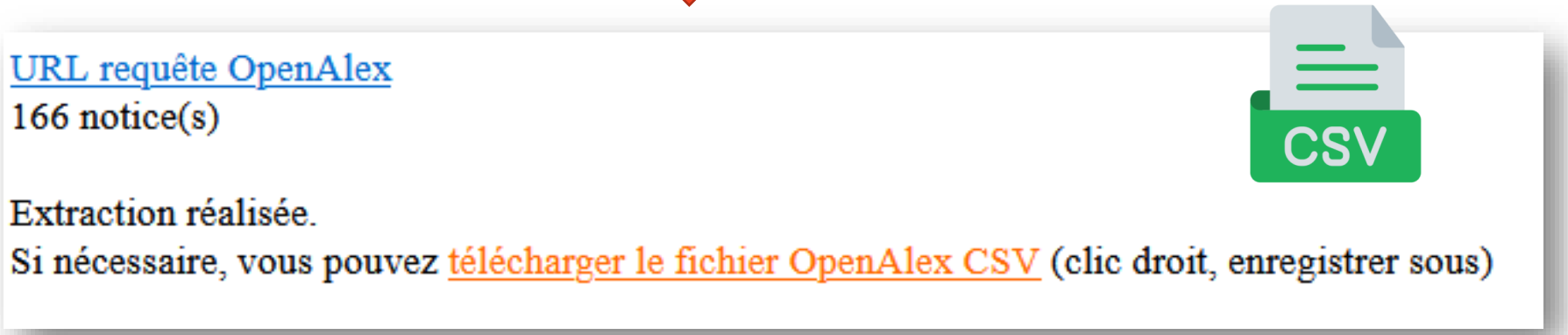
Etape 2 : Dédoublonner + générer un TEI HAL

OverHAL : générer le fichier CSV de la requête API



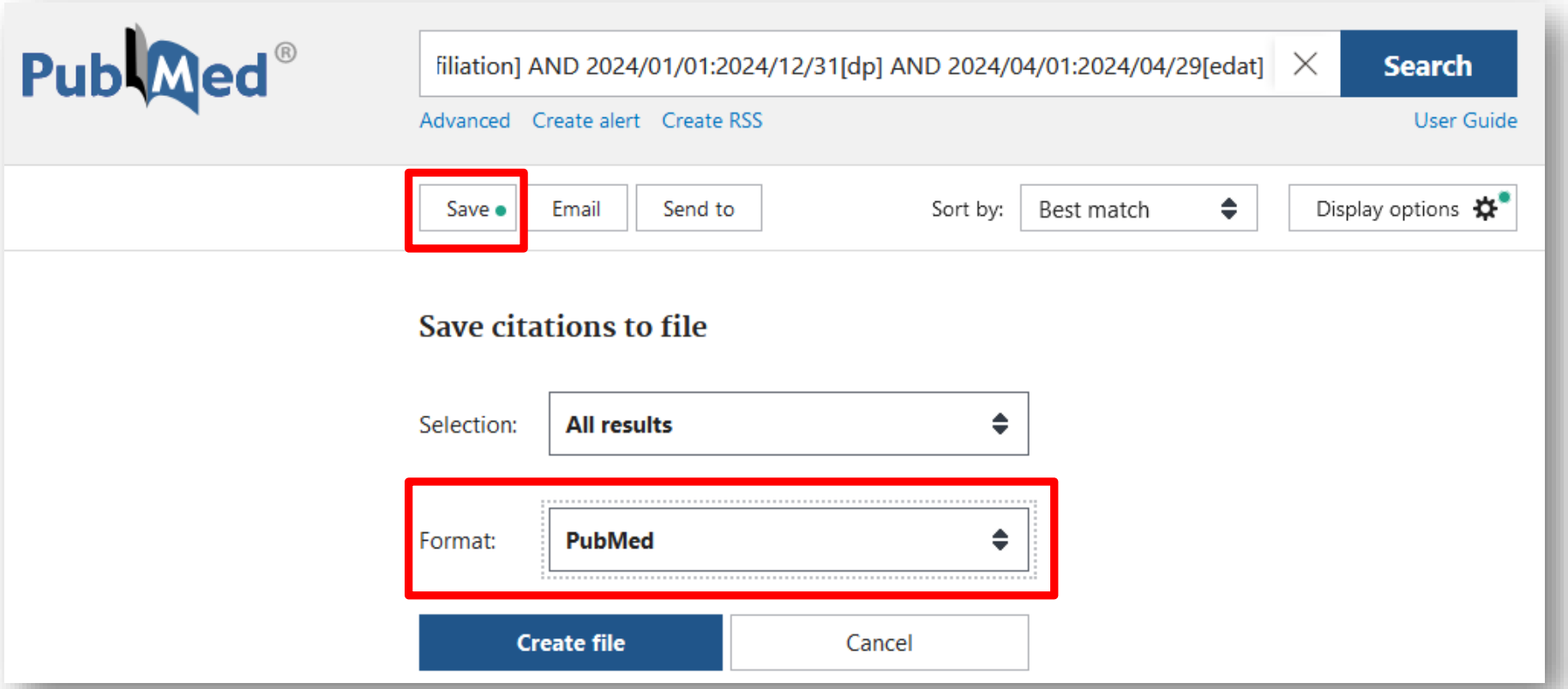
https://ccsdrennesvip.ccsd.cnrs.fr/OverHAL_OpenAlex_construct_import.php





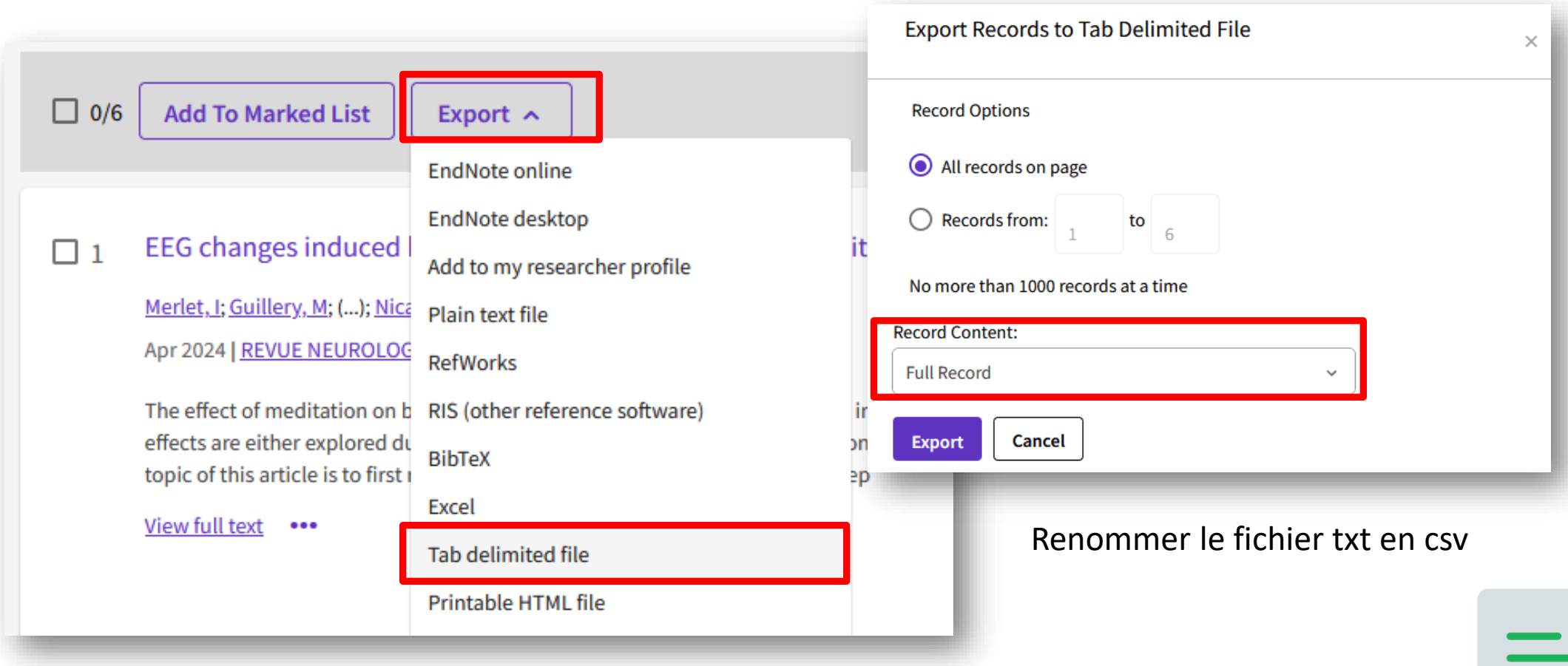
Université de Rennes

OverHAL : autres sources (Pubmed)



Une seule source à la fois (OpenAlex OU Pubmed OU WoS)

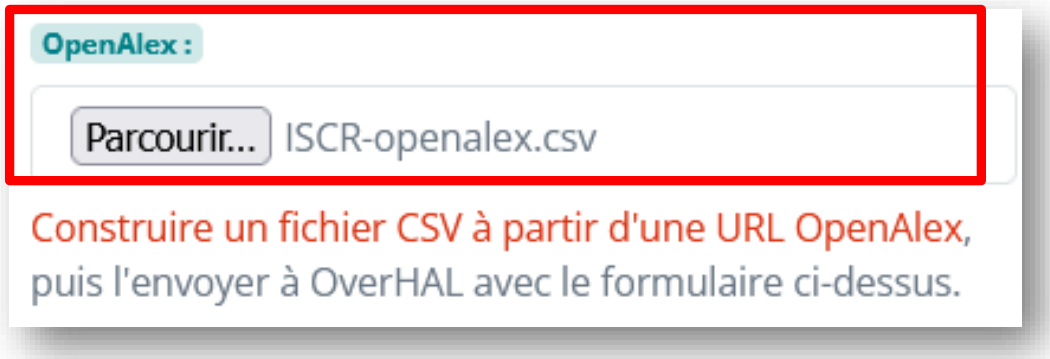
OverHAL : autres sources (Web of Science)



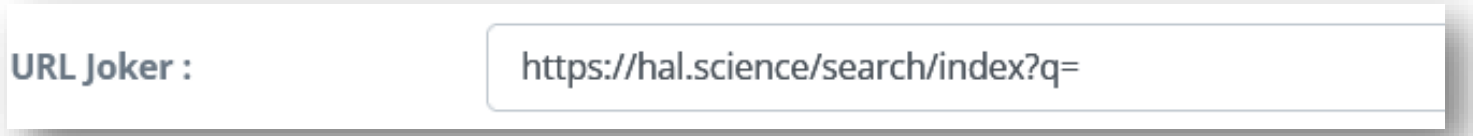
Renommer le fichier txt en csv

Une seule source à la fois (OpenAlex OU Pubmed OU WoS)

OverHAL : Import du fichier CSV



Attention à la **taille** du fichier !
Si trop de résultats, faire une requête sur une période de publication plus resserrée



OverHAL : résultats

Références de OpenAlex non trouvées dans HAL - *Retour aux résultats*

Attention, il est possible que la référence soit présente dans HAL mais qu'elle n'ait pas été trouvée en raison d'une légère différence dans le titre.
Par ailleurs, les notices avec plus de 50 auteurs ne seront pas prises en compte dans l'export TEI et apparaîtront ~~barrées~~.

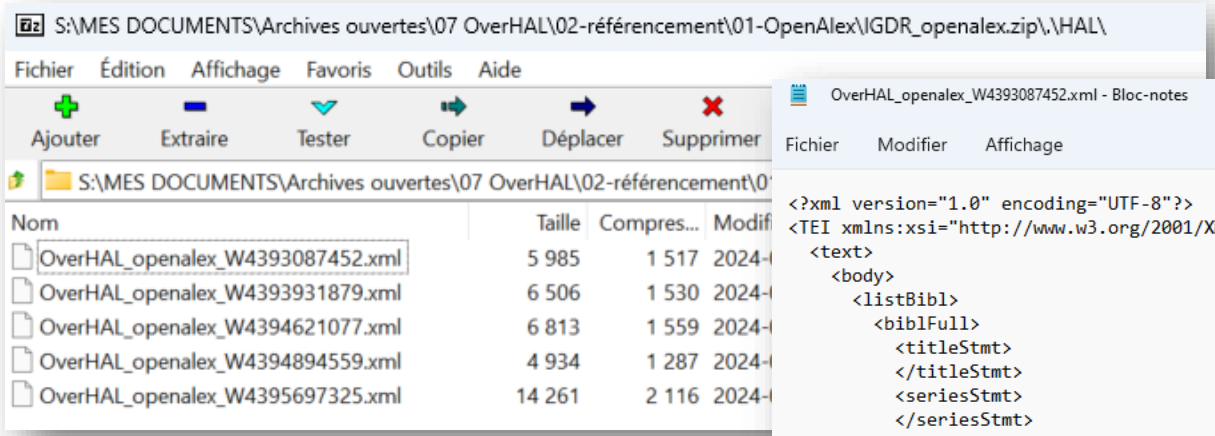
	Références	DOI	Source	Joker
1	Marie Cordier, Daniel S. Müller, Marc Devillard (2024) HMPA-Free Synthesis of TMS-Substituted Alkynes - *Organometallics* - doi: https://doi.org/10.1021/acs.organomet.4c00091			
2	Gabriel Dunes, Peter M. Chapple, Samia Kahlal, Thierry Roisnel, Jean-Francois Carpentier, Jean-Yves Saillard, Yann Sarazin (2024) Barium phosphidoboranes and related calcium complexes - *Dalton transactions* - doi: https://doi.org/10.1039/d4dt00487f			
3	Mateja Seničar, Benoît Roubinet, Pierre Lafite, Laurent Legentil, Vincent Ferrières, Ludovic Landemarre, Richard Daniellou (2024) Galf-Specific Neolectins: Towards Promising Diagnostic Tools - *International journal of molecular sciences* - doi: https://doi.org/10.3390/ijms25094826			
4	Hadrien Flichot, Annika Sickinger, Jules Brom, Bertrand Lefeuvre, Vincent Dorcet, Thierry Guizouarn, Olivier Cador, Boris Le Guennic, laurent.micouin@u-paris.fr laurent.micouin@u-paris.fr, Olivier Maury, erica.benedetti@u-paris.fr erica.benedetti@u-paris.fr, Fabrice Pointillart (2024) Magneto-structural correlation in lanthanide luminescent [2.2]Paracyclophane-based Single-Molecule Magnets - *Dalton transactions* - doi: https://doi.org/10.1039/d4dt00536h			

Université de Rennes

Exporter les résultats OpenAlex au format TEI

ISCR_openalex.zip

Fichier Zip



A ce stade, OverHAL a déjà calculé certaines métadonnées (affiliations) sous forme de chaîne de caractères (ex : « UMR6226 »)

OverHAL_openalex_W4393087452.xml - Bloc-notes

Fichier Modifier Affichage

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.tei-c.org/ns/1.0" xmlns:hal="http://hal.archives-ouvertes.fr/" xsi:schemaLo
  <text>
    <body>
      <listBibl>
        <biblFull>
          <titleStmt>
          </titleStmt>
          <seriesStmt>
          </seriesStmt>
          <notesStmt>
            <note type="audience" n="2"/>
            <note type="popular" n="0">No</note>
            <note type="peer" n="1">Yes</note>
          </notesStmt>
          <sourceDesc>
            <biblStruct>
              <analytic>
                <title xml:lang="en">Author response: ESCRT-III-dependent adhesive and mechanical changes are triggered by a mechanism detecting alterat
                <author role="aut">
                  <persName>
                    <forename type="first">Thomas</forename>
                    <surname>Esmangart de Bournonville</surname>
                  </persName>
                  <idno type="ORCID">https://orcid.org/0000-0001-6012-1726</idno>
                  <rawAffs>Global Health Institute, School of Life Science, Ecole Polytechnique Fédérale de Lausanne (EPFL) - Univ Rennes, CNRS, IGDR (I
                  <affiliation ref="#localStruct-Aff1"/>
                  <affiliation ref="#localStruct-Aff2"/>
                </author>
                <author role="aut">
                  <persName>
                    <forename type="first">Mariusz</forename>
                    <surname>K Jaglarz</surname>
                  </persName>
                  <idno type="ORCID">https://orcid.org/0000-0002-1606-8339</idno>
                  <rawAffs>Department of Developmental Biology and Invertebrate Morphology, Institute of Zoology and Biomedical Research, Jagiellonian U
                  <affiliation ref="#localStruct-Aff3"/>
                </author>
                <author role="aut">
                  <persName>
```

Ln 1, Col 1 100% Windows (CRLF) UTF-8

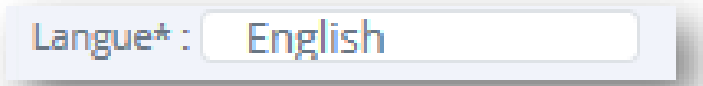
Université de Rennes

Etape 3 : Déposer dans HAL

Zip2HAL : points de vigilance

- Dans Zip2HAL, l'utilisateur dépose notice par notice, l'une après l'autre. On ne traite la notice suivante qu'après avoir déposé dans HAL la précédente.

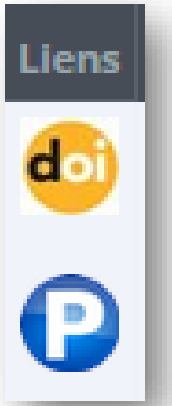
- Si la langue du document doit être modifiée, cette modification doit être réalisée en amont de toutes les autres.



- Toute modification du formulaire de saisie modifie en direct le fichier TEI.

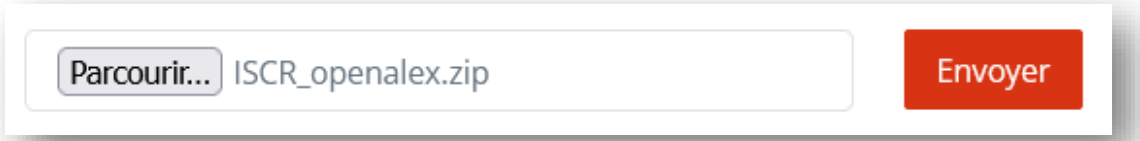
- IMPORTANT : n'effacez jamais votre saisie avec la souris ou votre clavier. Utilisez toujours la poubelle. Sans quoi l'information risque de ne pas être effacée du TEI et bloquer votre dépôt.

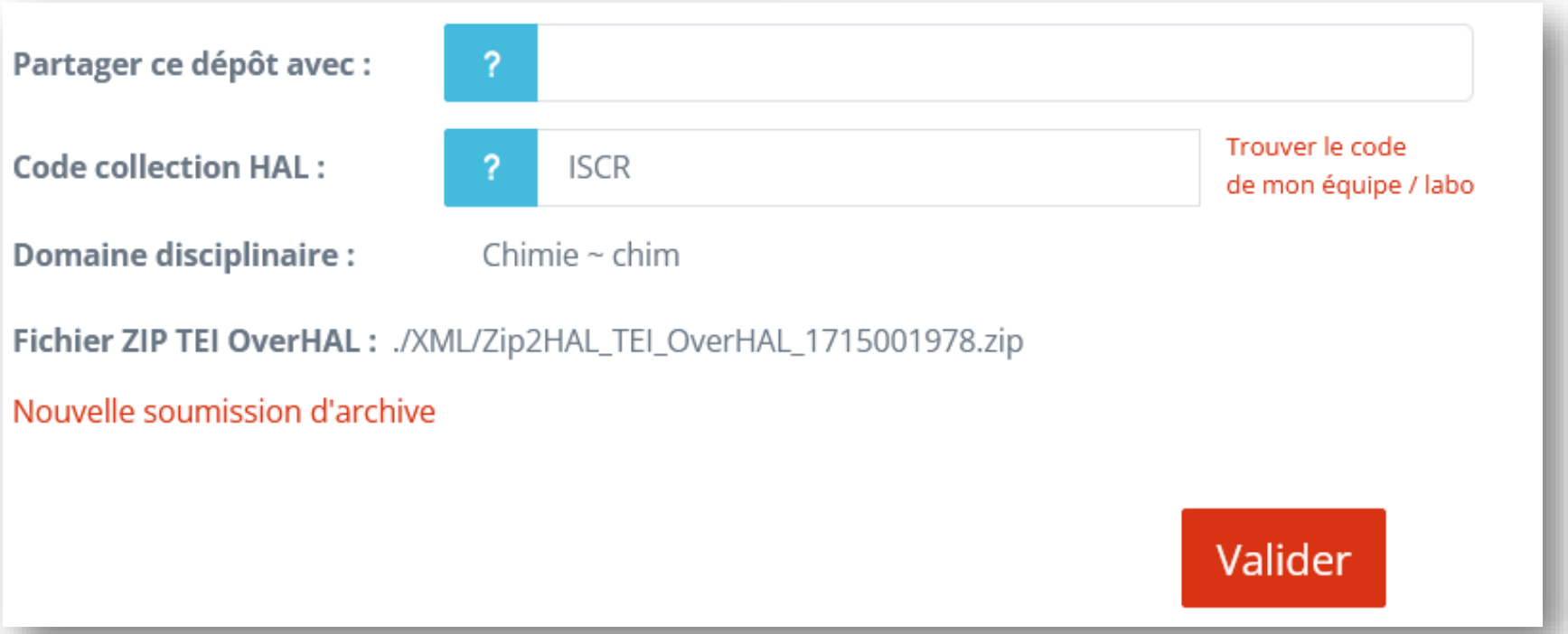
- Cliquez sur le lien DOI ou Pubmed pour comparer les données Zip2HAL et les données de l'article afin d'éviter toute erreur.



- Les données obligatoires sont indiquées par un astérisque *

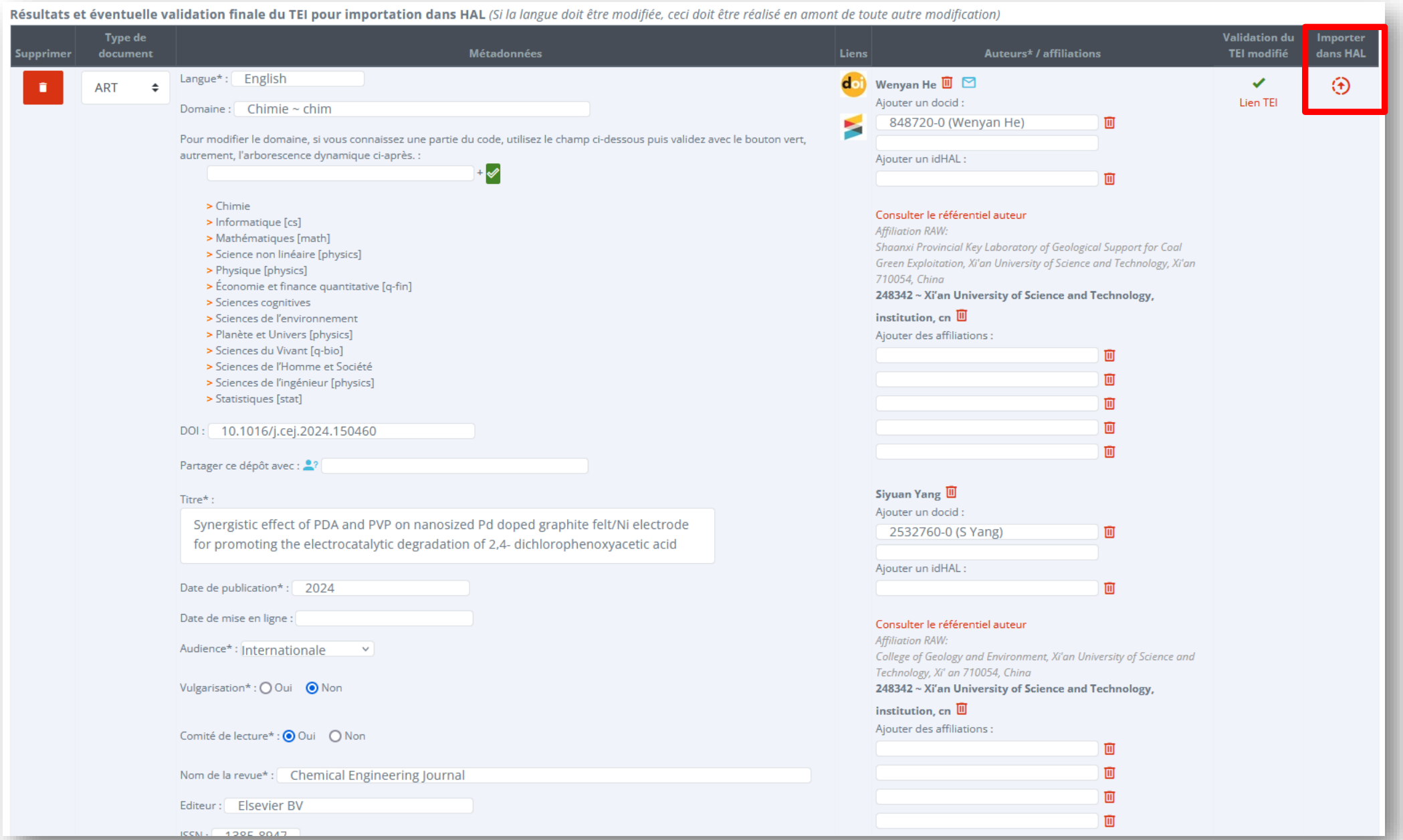
Zip2HAL : Chargement du fichier zip





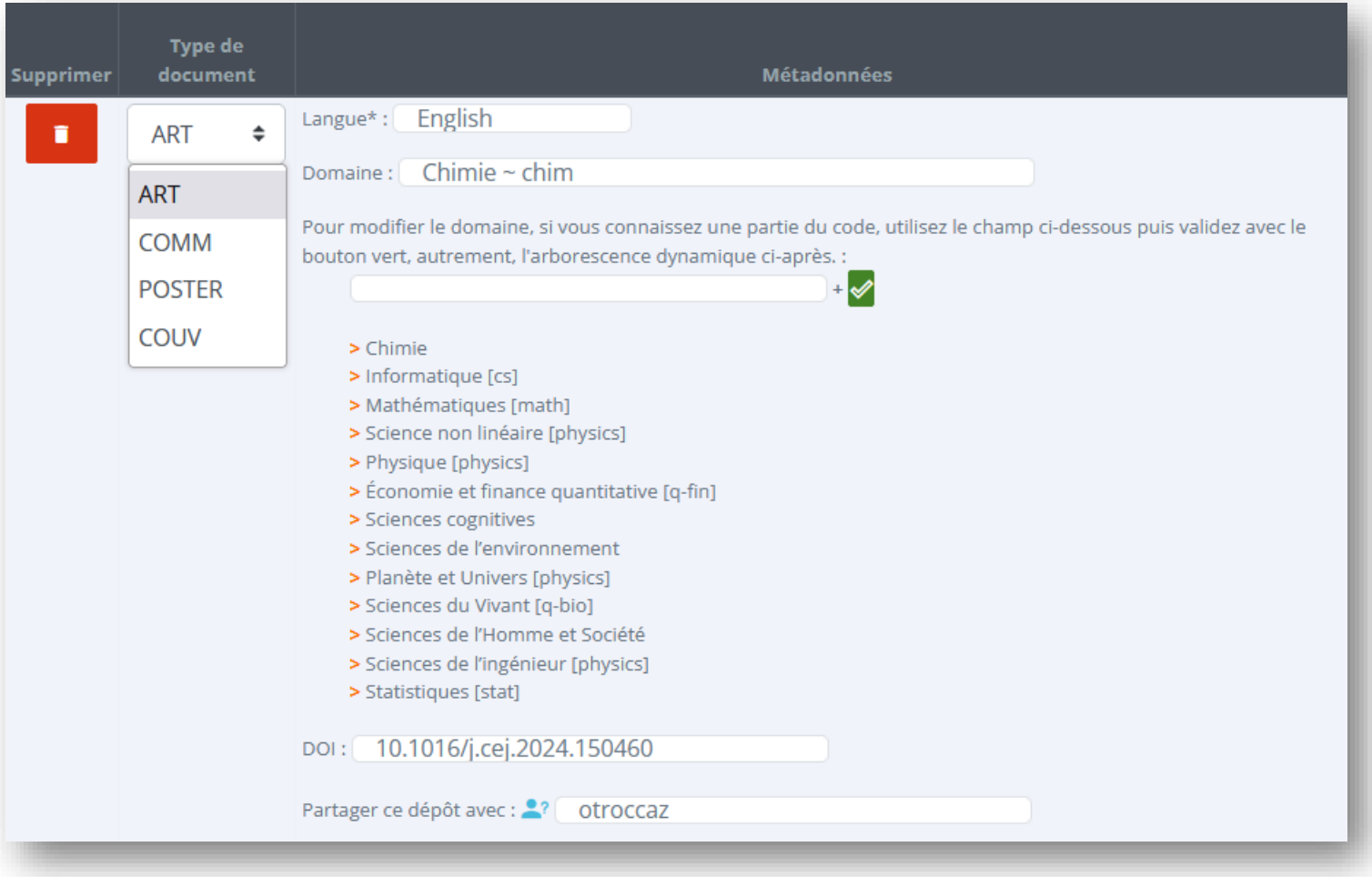
Attendre que le script ait terminé l'analyse du fichier avant de passer à l'étape suivante !

Zip2HAL : interface utilisateur (vue d'ensemble)



1. Lien DOI
2. Métadonnées
3. Auteurs et affiliations
4. Import dans HAL

Zip2HAL : métadonnées



Type de document : peu fiable entre ART et COMM / POSTER

Langue : la modification doit être faite en amont de tout autre changement (génère des champs spécifiques)

Domaine : sélectionner un domaine dans l'arborescence ou entrer le code et valider (ex : « sdv.ca »)

Partage du dépôt : ne fonctionne qu'avec un seul login

Champs obligatoires signalés par un astérisque *

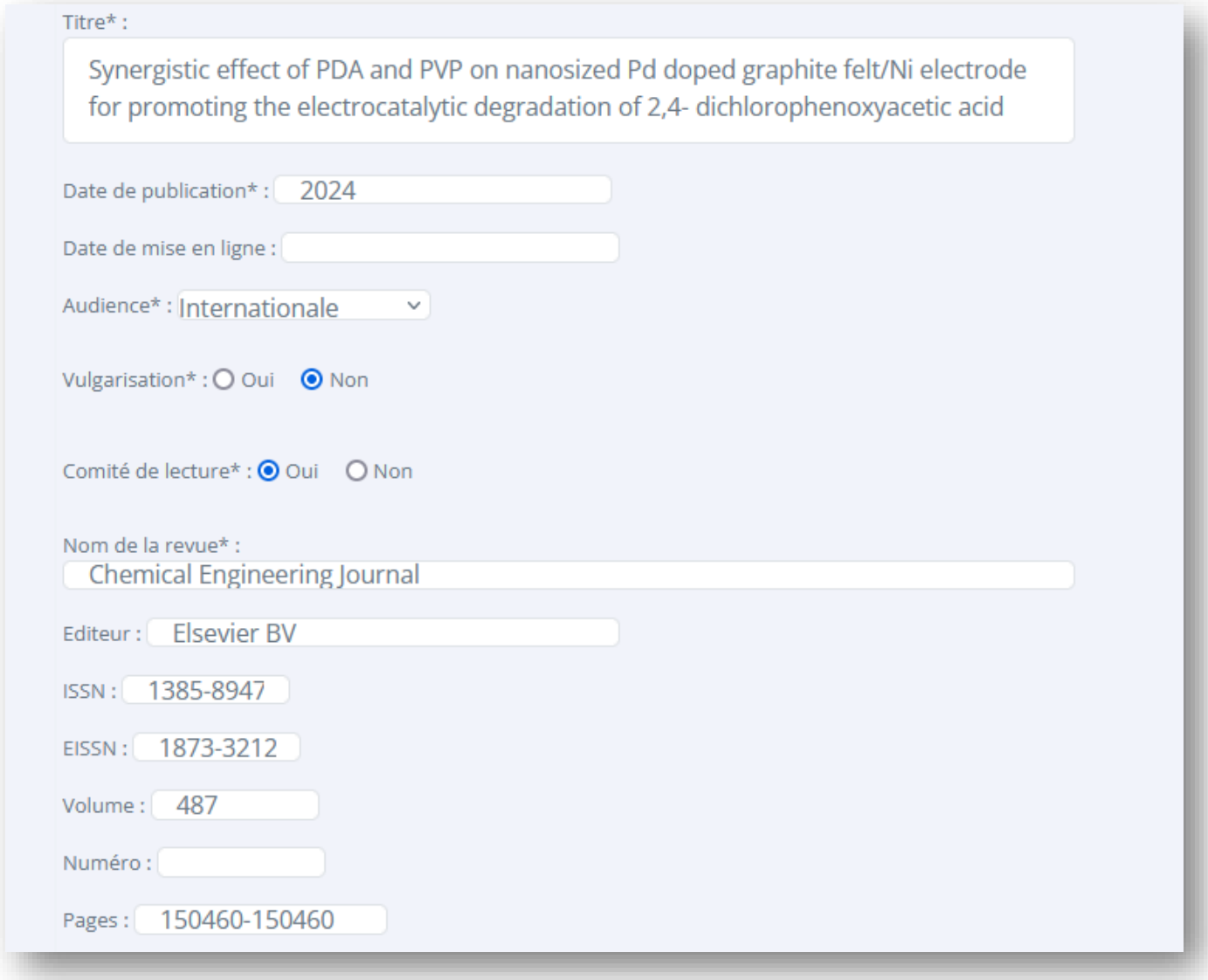
Zip2HAL : métadonnées

Titre : supprimer les éventuelles balises html (ex :)

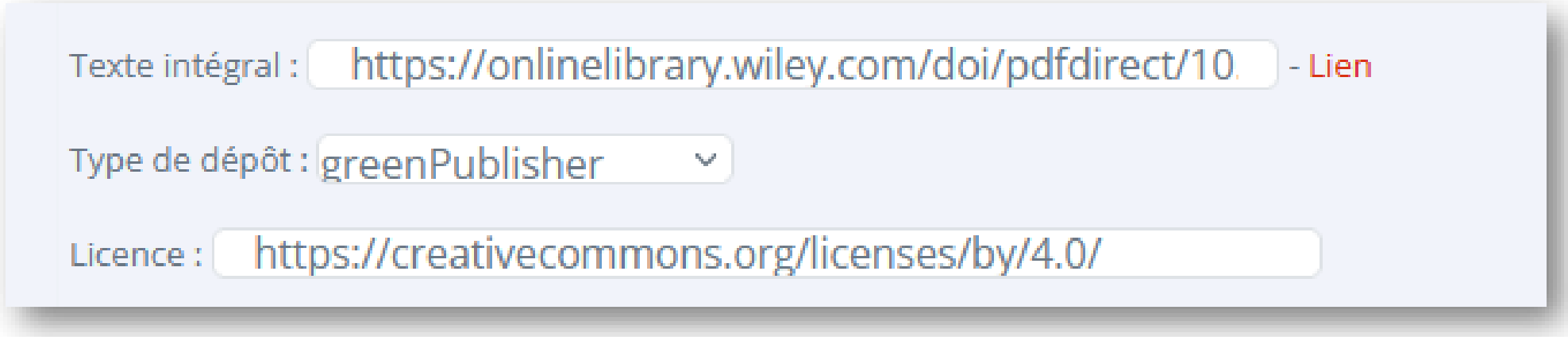
Audience : par défaut, « internationale » si langue = EN

ISSN/EISSN : il peut être parfois nécessaire de modifier cette donnée dans le formulaire ou dans AuréHAL (si erreur ISSN au dépôt)

Pagination : il s'agit parfois d'un numéro, souvent redoublé dans OpenAlex (dans cet exemple : 150460)



Zip2HAL : métadonnées



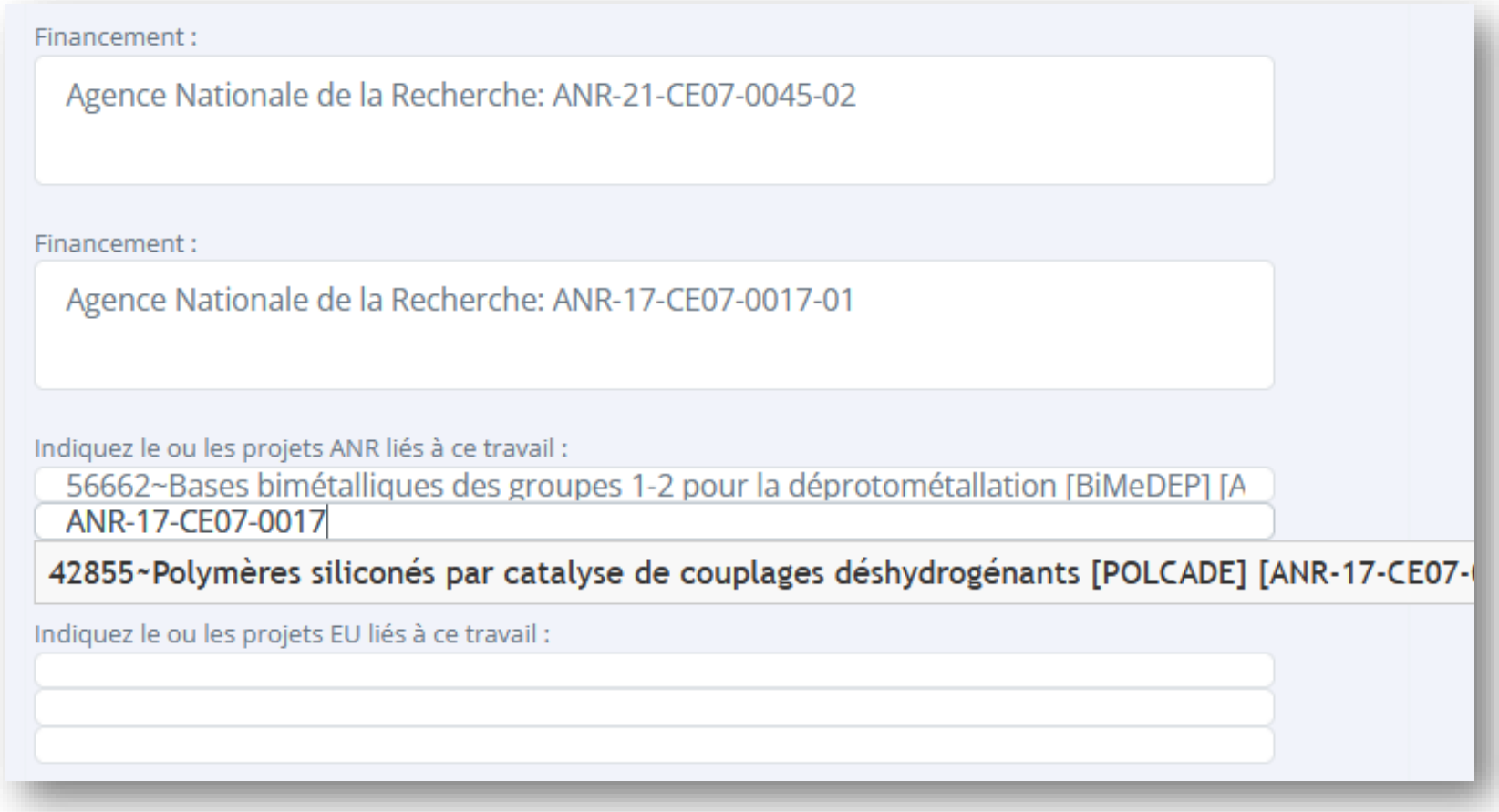
Le lien du fichier PDF est remonté de Unpaywall (à vérifier)

Le lien n'aboutit pas toujours à un dépôt en texte intégral dans HAL (dépend des éditeurs)

Type de dépôt = greenPublisher : correspond au champ HAL « J'ai vérifié que cet éditeur autorise le dépôt du PDF éditeur »

Attention : la licence CC n'est pas toujours précisée dans OpenAlex. Vérifier qu'il ne s'agit pas d'un article « free » mais sous copyright chez l'éditeur (dans ce cas, effacer le lien)

Zip2HAL : métadonnées



Données financement parfois absentes ou incomplètes dans openAlex (à copier-coller). « Acknowledgements » ou « Funding »

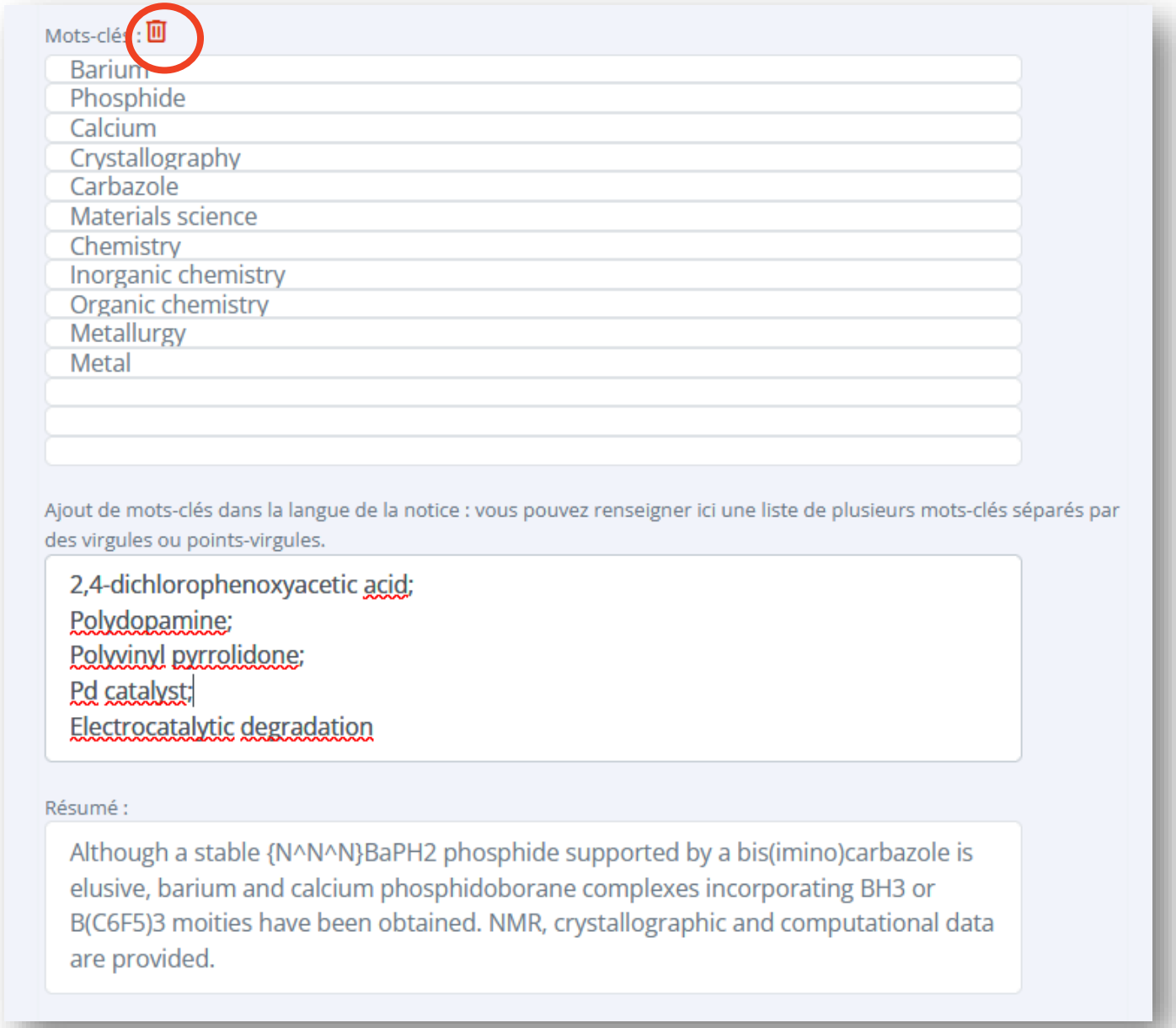
Ajouter les codes décision ANR dans les champs spécifiques (recherche dans le référentiel AuréHAL)

Zip2HAL : métadonnées

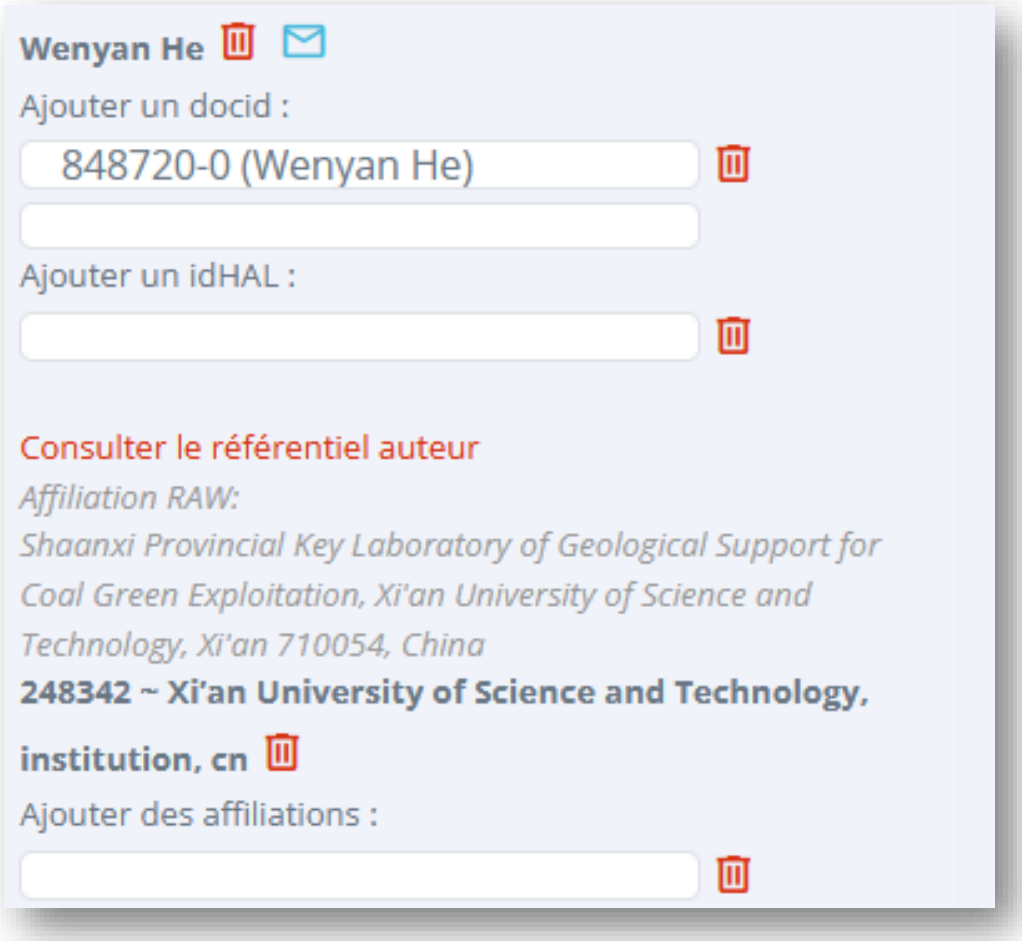
Mots-clés : ceux d'openalex ne sont pas toujours fiables et ne proviennent pas de l'éditeur.

Ajout de mots-clés : données ajoutées par l'utilisateur, par copier-coller (séparateur virgule ou point-virgule)

Résumé : souvent absent ou incomplet dans OpenAlex, à remplacer par le résumé de l'éditeur (copier-coller)



Zip2HAL : auteurs et affiliations



Auteur : le script a trouvé dans AuréHAL une forme auteur sans identifiant associé (email, ORCID, etc) : 84720 (ID Forme) et 0 (ID Auteur) – **attention :** toujours bien vérifier que le nom ajouté dans le champ correspond bien au nom affiché en gras, il peut y avoir de fausses correspondances (l'algorithme n'est pas infaillible)

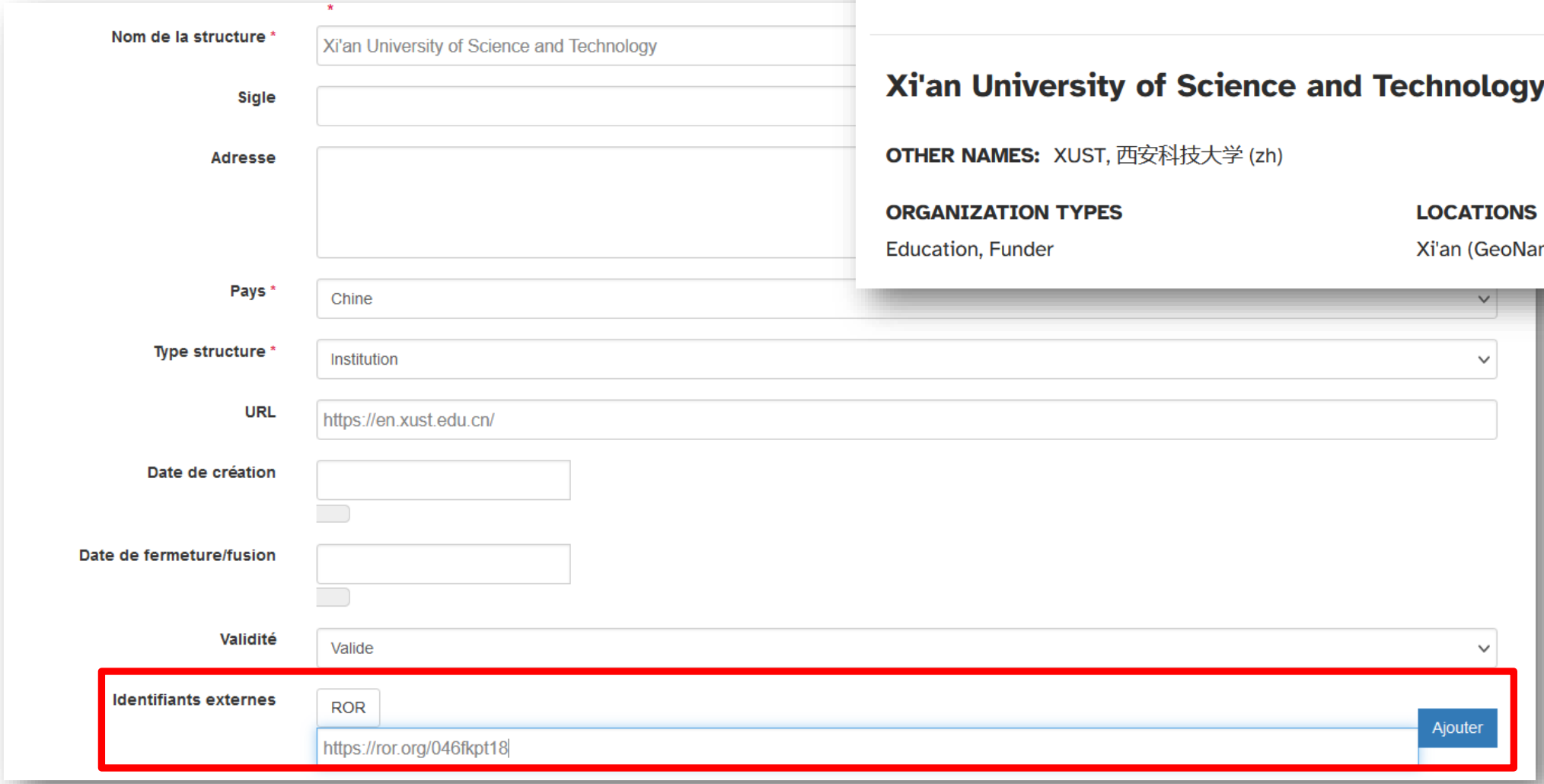
L'icône enveloppe signale un auteur correspondant. On peut l'activer le cas échéant (en grisé).

Affiliation : comparer la source (en gris) issue d'OpenAlex et la forme calculée (en gras) pour **vérifier que l'affiliation retournée est correcte**. Il manque l'encoche verte pour signaler que le calcul a été fait sur le ROR > le ROR est probablement manquant dans AuréHAL > l'ajouter : à la prochaine itération, le script retrouvera le ROR

Zip2HAL : auteurs et affiliations

auréHAL

https://aurehal.archives-ouvertes.fr/structure/



Ce peut être l'occasion de renseigner le ROR des institutions étrangères dans HAL => améliore le calcul des affiliations

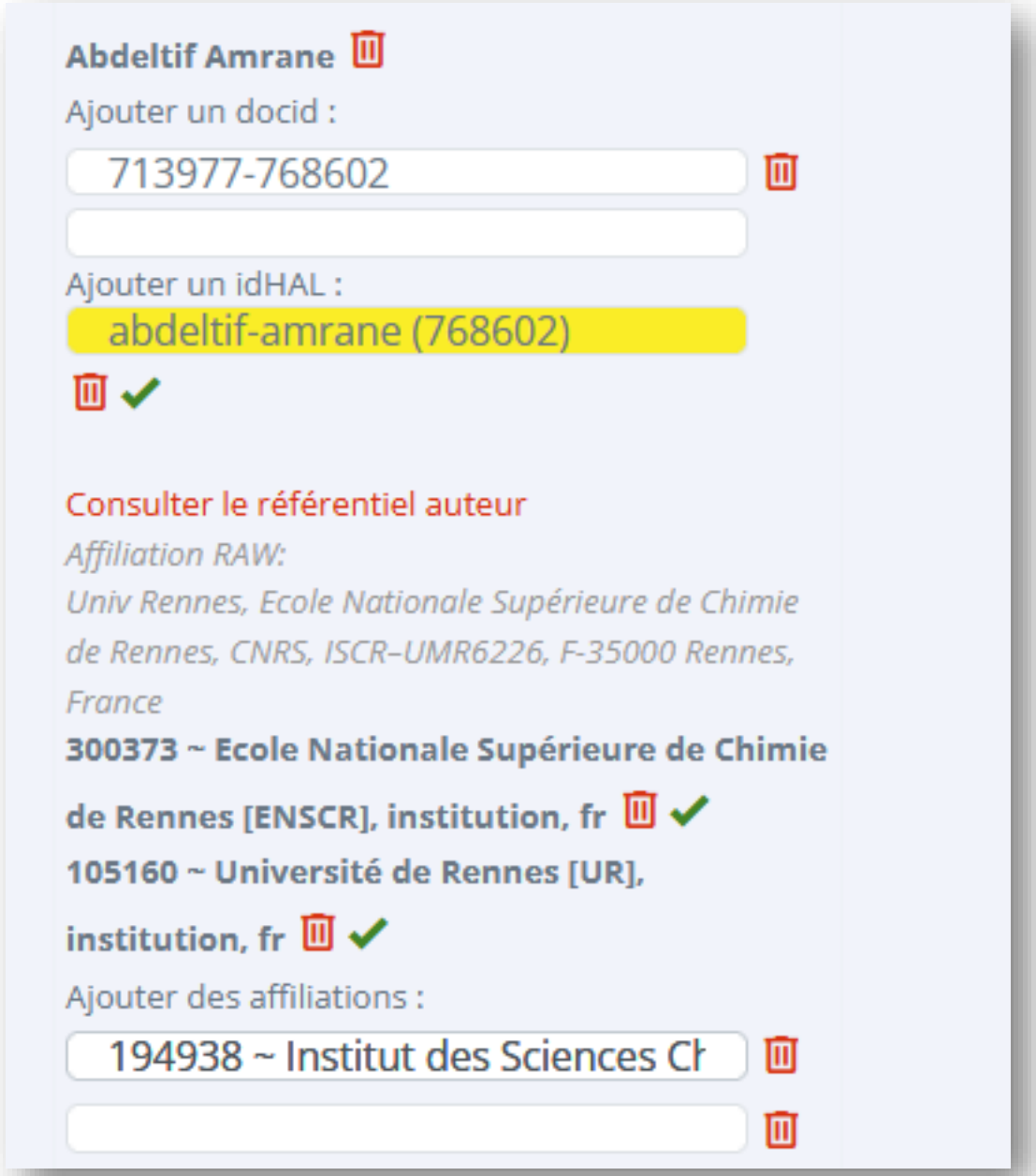
Zip2HAL : auteurs et affiliations

L'idHAL a été remonté grâce à l'ORCID (encoche verte)

Les ROR des affiliations ont été trouvés dans AuréHAL (encoche verte)

Une affiliation manquante (UMR6226), pour laquelle le ROR existe, mais que l'algorithme d'OpenAlex n'a pas su calculer

Importance de la charte de signature dans les articles : le nom de l'unité n'est pas renseigné, l'acronyme et le code UMR sont difficilement identifiables pour l'algorithme (une seule chaîne « ISCR-UMR6226 »). Comparer :



Zip2HAL : auteurs et affiliations

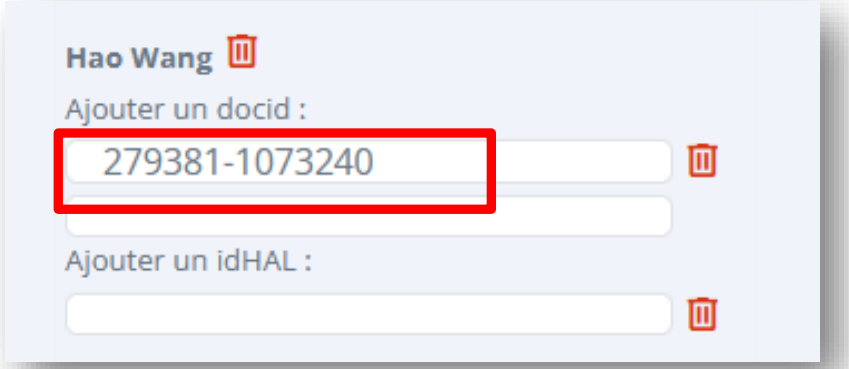
Ce peut être l'occasion de vérifier les informations des auteurs ayant un idHAL pour repérer les dépôts non associés à l'idHAL

Zip2HAL : auteurs et affiliations

Le docid auteur retourné est faux : il faut le supprimer (icône poubelle) et rechercher le bon docid dans AuréHAL – s'il existe

ID Forme	ID Auteur	idHAL	Nom	Prénom	Second Prénom	Domaine email
279381	1073240		Hao	Wang		stu.ecnu.edu.cn

Il suffit ensuite de copier-coller le docid AuréHAL

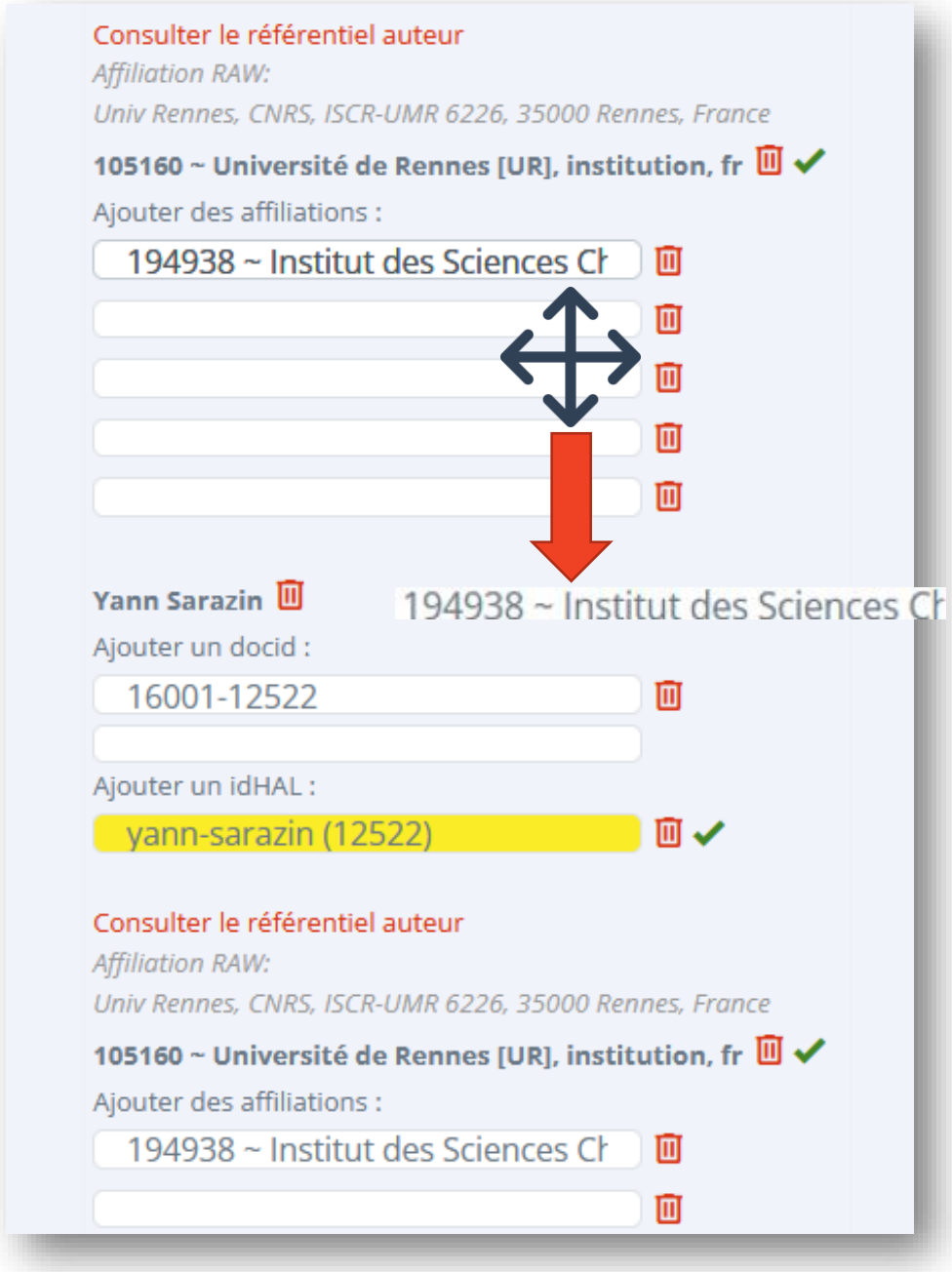


Zip2HAL : auteurs et affiliations

Glisser-déposer avec la souris pour ajouter la même affiliation à plusieurs auteurs

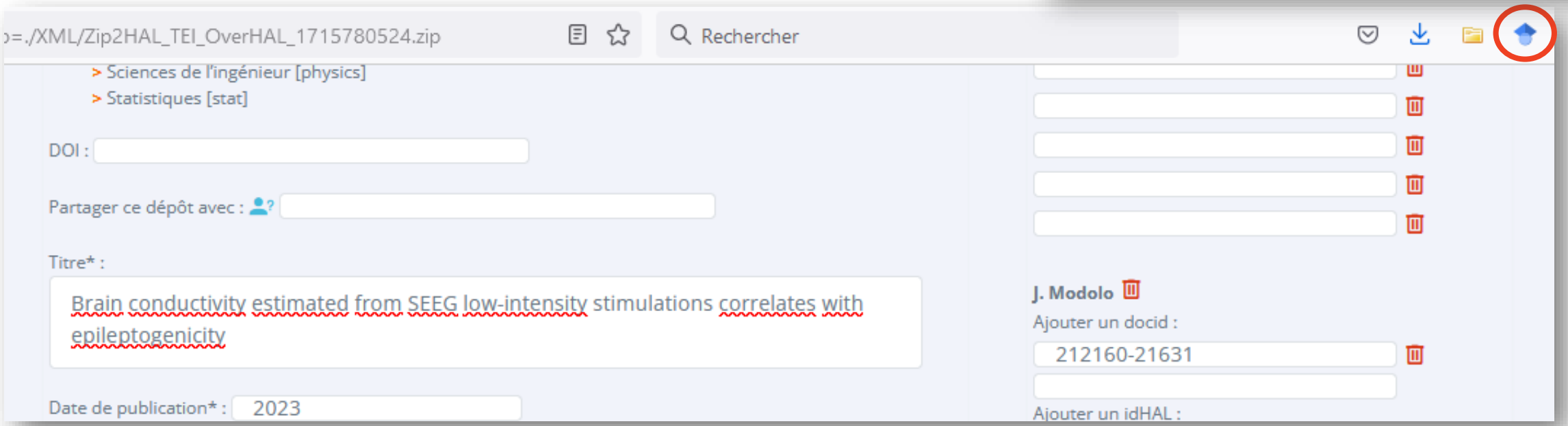
Ne pas lâcher la souris lorsqu'on ajoute une affiliation supplémentaire dans le formulaire (saisie texte) – risque d'inscrire une donnée erronée dans le XML

Toujours utiliser le bouton « poubelle » pour supprimer une donnée



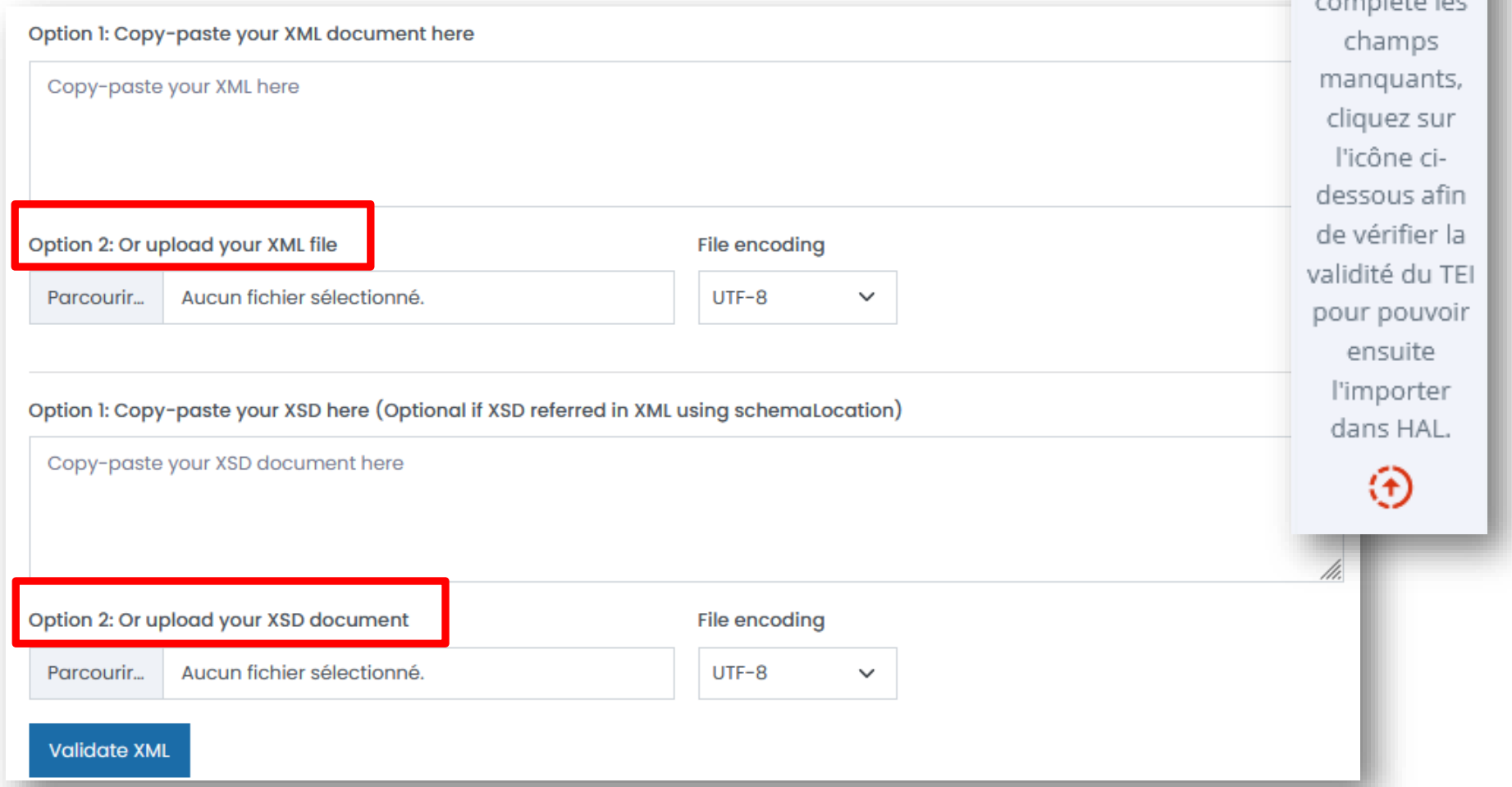
Zip2HAL : Absence de DOI

Faire une recherche dans l'extension Google Scholar pour retrouver un éventuel DOI et s'assurer qu'on ne crée pas un doublon



Zip2HAL : validation du TEI modifié (utilisation avancée)

https://www.freeformatter.com/xml-validator-xsd.html#



FreeFormatter permet de repérer les erreurs qui empêchent la validation du TEI

Document XSD : https://hal.science/documents/aofr.xsd

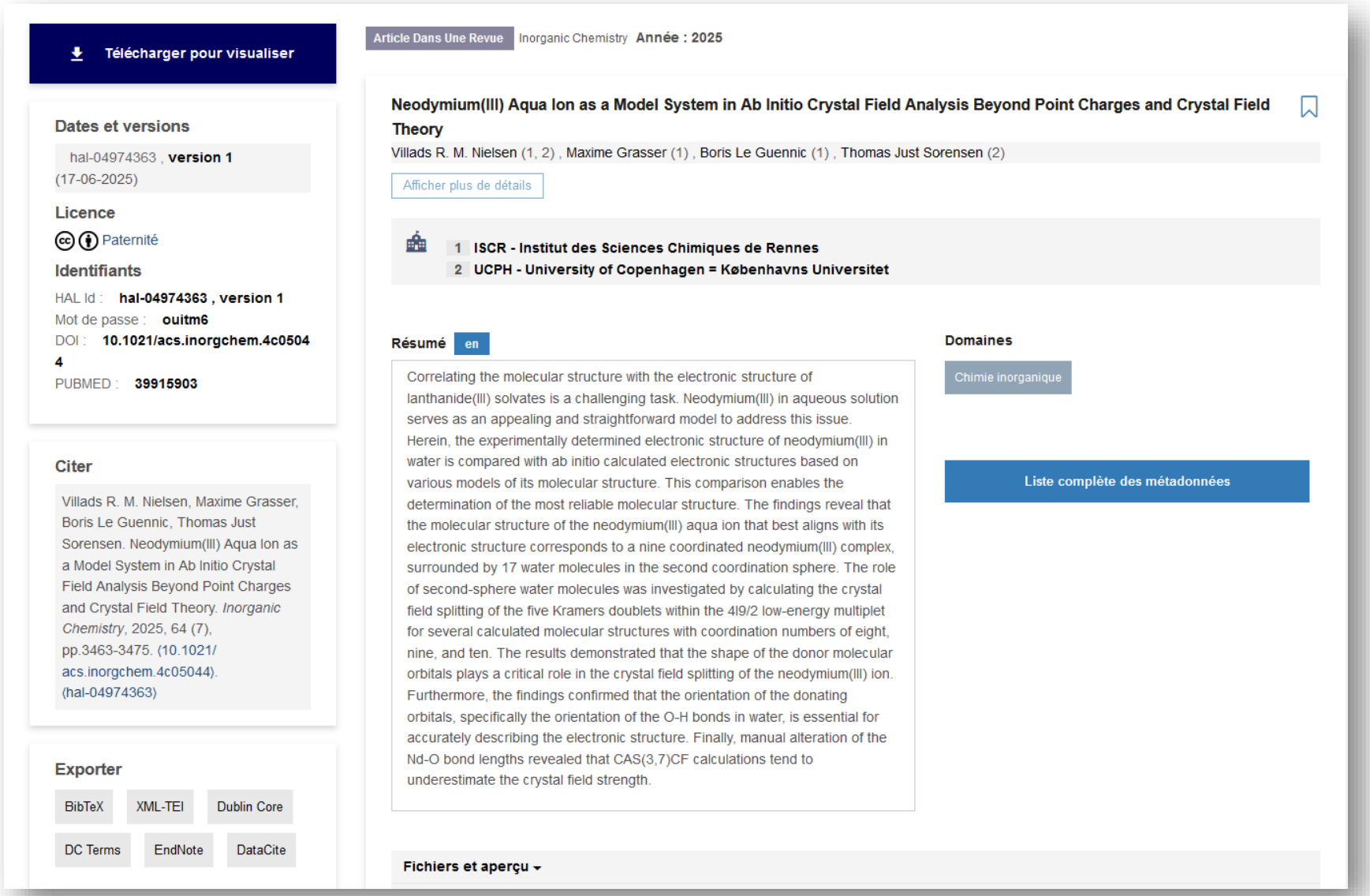
Zip2HAL : Lien TEI et dépôt

Pas d'import par lot, on importe la notice une fois les métadonnées vérifiées

```
−<TEI xsi:schemaLocation="http://www.tei-c.org/ns/1.0 http://api.archives-ouvertes.fr/documents/aofr-sword.xsd">
  −<text>
    −<body>
      −<listBibl>
        −<biblFull>
          <titleStmt> </titleStmt>
          −<editionStmt>
            −<edition>
              <ref type="file" subtype="greenPublisher" n="1" target="https://pubs.rsc.org/en/content/articlepdf/2024/su/d3su00470h"/>
            </edition>
          </editionStmt>
          −<publicationStmt>
            −<availability>
              <licence target="https://creativecommons.org/licenses/by/4.0/"/>
            </availability>
          </publicationStmt>
          −<seriesStmt>
            <idno type="stamp" n="UNIV-RENNES1"/>
          </seriesStmt>
```

Exemple de fichier TEI envoyé par Zip2HAL via l'API SWORD

La notice HAL est créée (nouvel onglet)



Télécharger pour visualiser

Dates et versions

hal-04974363 , **version 1**
(17-06-2025)

Licence

Paternité

Identifiants

HAL Id : **hal-04974363 , version 1**
Mot de passe : **ouitm6**
DOI : **10.1021/acs.inorgchem.4c05044**
PUBMED : **39915903**

Citer

Villads R. M. Nielsen, Maxime Grasser, Boris Le Guennic, Thomas Just Sorensen. Neodymium(III) Aqua Ion as a Model System in Ab Initio Crystal Field Analysis Beyond Point Charges and Crystal Field Theory. *Inorganic Chemistry*, 2025, 64 (7), pp.3463-3475. (10.1021/acs.inorgchem.4c05044). (hal-04974363)

Exporter

BibTeX | XML-TEI | Dublin Core | DC Terms | EndNote | DataCite

Article Dans Une Revue Inorganic Chemistry **Année : 2025**

Neodymium(III) Aqua Ion as a Model System in Ab Initio Crystal Field Analysis Beyond Point Charges and Crystal Field Theory

Villads R. M. Nielsen (1, 2) , Maxime Grasser (1) , Boris Le Guennic (1) , Thomas Just Sorensen (2)

Afficher plus de détails

1 **ISCR - Institut des Sciences Chimiques de Rennes**
2 **UCPH - University of Copenhagen = Københavns Universitet**

Résumé en

Correlating the molecular structure with the electronic structure of lanthanide(III) solvates is a challenging task. Neodymium(III) in aqueous solution serves as an appealing and straightforward model to address this issue. Herein, the experimentally determined electronic structure of neodymium(III) in water is compared with ab initio calculated electronic structures based on various models of its molecular structure. This comparison enables the determination of the most reliable molecular structure. The findings reveal that the molecular structure of the neodymium(III) aqua ion that best aligns with its electronic structure corresponds to a nine coordinated neodymium(III) complex, surrounded by 17 water molecules in the second coordination sphere. The role of second-sphere water molecules was investigated by calculating the crystal field splitting of the five Kramers doublets within the 4I9/2 low-energy multiplet for several calculated molecular structures with coordination numbers of eight, nine, and ten. The results demonstrated that the shape of the donor molecular orbitals plays a critical role in the crystal field splitting of the neodymium(III) ion. Furthermore, the findings confirmed that the orientation of the donating orbitals, specifically the orientation of the O-H bonds in water, is essential for accurately describing the electronic structure. Finally, manual alteration of the Nd-O bond lengths revealed that CAS(3,7)CF calculations tend to underestimate the crystal field strength.

Domaines

Chimie inorganique

Liste complète des métadonnées

Fichiers et aperçu

Université de Rennes

Zip2HAL : erreurs Sword les plus courantes

Erreurs Sword	Solutions
Code retour http : 0 ERREUR : **Problème avec l'API sword**, contactez le support technique (erreur http=0)	Vérifier dans Mes dépôts : le dépôt est effectif malgré le message d'erreur
Code retour http : 400 ERREUR : Pb sword : Internal server error: **Firstname** (Pierre-olivier) **is invalid**	Supprimer l'auteur dont l'affiliation ou la forme empêche le dépôt et l'ajouter ensuite dans la notice HAL créée
Code retour http : 400 ERREUR : Pb sword : {"**duplicate-entry**":{"**hal-04575964**":{"doi":"1.0","pubmed":"1.0"}}}	Notice doublon non détectée par le script (titre différent, absence de DOI, preprint avec le même DOI…)
Code retour http : 400 ERREUR : Pb sword : {"mainmeta":{"**mainDomain**":{"**isEmpty**":"Cette valeur est obligatoire et ne peut \u00eatre vide"}}}	Le domaine disciplinaire n'a pas été renseigné
Code retour http : 400 ERREUR : Pb sword : **Could not open the XML** description: source XML incorrecte: DOMDocument::loadXML(): Start tag expected, '<' not found in Entity, line: 2	Bug sur le fichier XML pour les notices contenant beaucoup d'auteurs (non résolu à ce jour)

Echange : que vous inspire cet atelier ?

A Young Man Raising his Hand at a Business Conference – free to use

- Facile ou trop compliqué ?
- Pensez-vous utiliser ces outils ?
- « catalogage » et auto-archivage
- …

Soulevons le capot :
Comment travaille l'algorithme Zip2HAL ?

Zip2HAL

Zip2HAL : étapes du traitement automatique du fichier zip

Traitement du fichier OverHAL_openalex_W4392868848.xml

Recherche des doublons potentiels — Étape 1

URL requête API HAL - URL requête API HAL CRAC
Aucune notice trouvée dans HAL, ni HAL CRAC **donc, pas de doublon**

Recherche des idHAL et docid des auteurs — Étape 2

Recherche idHAL/docid
1 idHal et 9 docid trouvé(s)

Recherche des id structures des affiliations — Étape 3a

Calcul des affiliations
5 id structures des affiliations trouvé(s)

Recherche de la dernière affiliation associée avec HAL aux auteurs sans affiliation — Étape 3b

Calcul des dernières affiliations
0 affiliation(s) manquante(s) trouvée(s)

Recherche des docid auteur grâce aux affiliations éventuellement trouvées — Étape 3c

Recherche des docid
0 docid auteur trouvé(s)

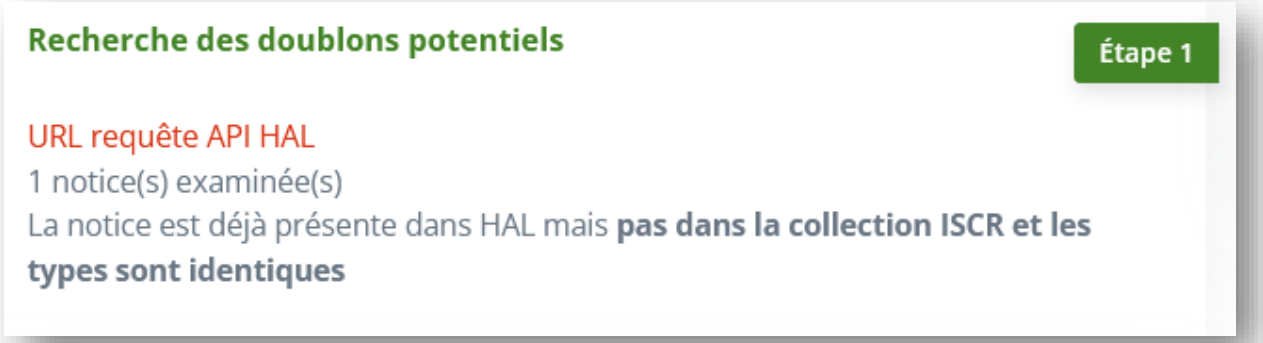
Recherche des idHAL auteur grâce aux docid auteur trouvés précédemment — Étape 3d

Recherche des idHAL
URL requête idHAL auteur
URL requête idHAL auteur
URL requête idHAL auteur
URL requête idHAL auteur
URL requête idHAL auteur
1 idHAL auteur trouvé(s)

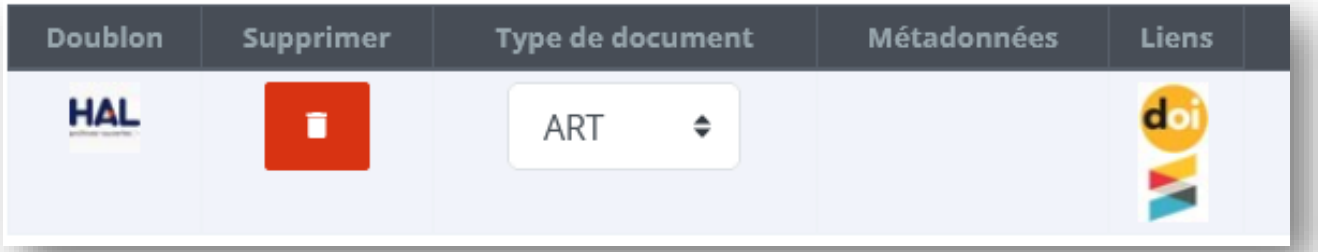
Attention : merci de vérifier très attentivement les affiliations et idHAL des auteurs calculés par l'algorithme (ci-dessous) : les résultats peuvent contenir des erreurs.

Université de Rennes

Zip2HAL : la notice est déjà dans HAL (à vérifier)

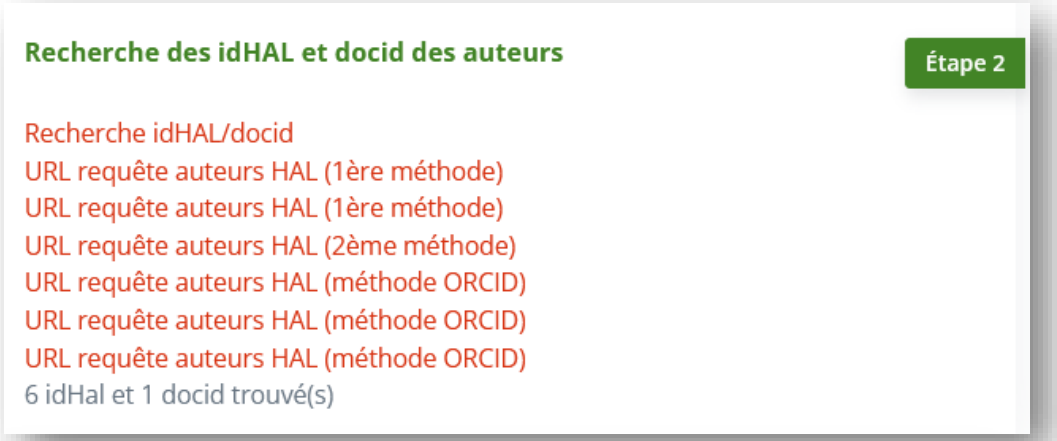


Recherche des doublons sur les *n* premiers mots du titre avec troncature * OU doi + date de publication (- / + 1 an) + à paraître >> peut manquer des doublons si légère différence dans le titre



La recherche de doublons est également faite sur les dépôts en modération (https://api.archives-ouvertes.fr/crac/hal/)

Zip2HAL : calcul des identifiants auteurs (référentiel HAL)




```
docid: "1329555-1143868",
fullName_s: "Hanna Oher",
valid_s: "PREFERRED",
emailDomain_s: [
    "univ-rennes1.fr"
],
idHal_i: 1143868,
idHal_s: "hanna-oher"
```


1ère méthode : recherche de l'id auteur à partir du nom complet (fullName_t), en privilégiant l'idHAL

2e méthode : même chose, avec une recherche exacte (fullName_sci)

Méthode ORCID, si la métadonnée est disponible dans le fichier OpenAlex


```
docid: "325-686",
fullName_s: "Valérie Vallet",
valid_s: "PREFERRED",
orcidId_s: [
    https://orcid.org/0000-0002-2202-3858
]
```

Zip2HAL : calcul des affiliations

Recherche de l'affiliation par le ROR, l'acronyme, le nom ou le code UMR


```
docid: "194938",
ror_s: [
    https://ror.org/00adwkx90
],
name_s: "Institut des Sciences Chimiques de Rennes",
acronym_s: "ISCR",
country_s: "fr",
type_s: "laboratory",
valid_s: "VALID"
```

Zip2HAL : calculs supplémentaires idHAL, affiliations,

Recherche de la dernière affiliation associée avec HAL aux auteurs sans affiliation — Étape 3b

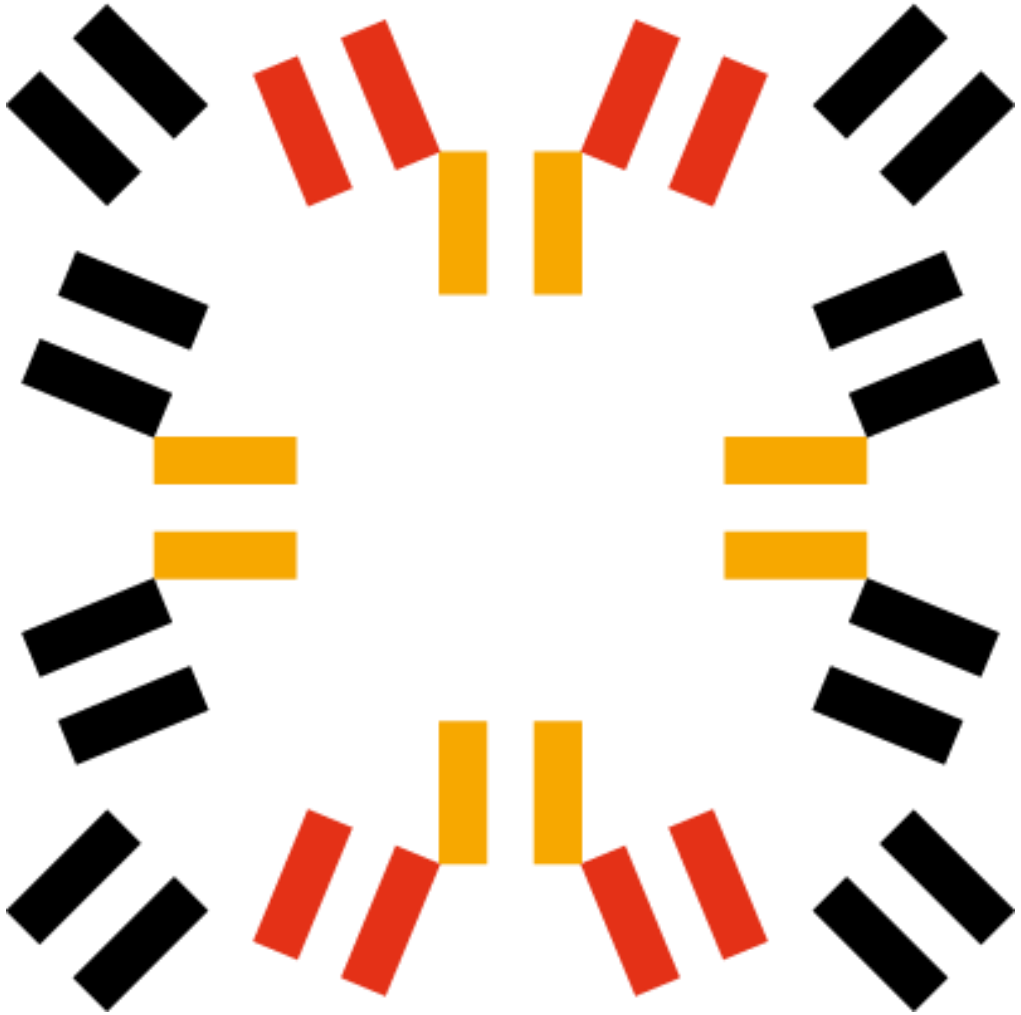
Calcul des dernières affiliations
URL requête auteur structure HAL
URL requête auteur structure HAL
URL requête auteur structure HAL
URL requête auteur structure HAL
URL requête test validité affiliation trouvée
1 affiliation(s) manquante(s) trouvée(s)

Recherche des docid auteur grâce aux affiliations éventuellement trouvées — Étape 3c

Recherche des docid
0 docid auteur trouvé(s)

Recherche des idHAL auteur grâce aux docid auteur trouvés précédemment — Étape 3d

Recherche des idHAL
URL requête idHAL auteur
URL requête idHAL auteur
URL requête idHAL auteur
1 idHAL auteur trouvé(s)



Merci de votre attention

Laurent.jonchere@univ-rennes.fr

Université de Rennes