

OverHAL – Manuel d'utilisation

OverHAL permet de comparer HAL et des listes de publications (Scopus, Web of Science, Zotero, etc), à partir d'un script PHP créé par [Philippe Gambette](#), repris et modifié par Olivier Troccaz (ECOBIO - OSUR) pour l'Université de Rennes 1.

A l'université de Rennes 1, OverHAL est utilisé pour :

- **Référencer des publications dans HAL** : le script permet de générer un fichier TEI-HAL ou Bibtex pour X2HAL
- **Gérer le dépôt du texte intégral** : envoi de mails aux auteurs à partir d'un export Zotero

Installation

Téléchargez le dossier zip sur la [page du wiki HAL](#) (OverHAL) et suivez les instructions de la notice d'installation incluse dans le zip.

Lien à utiliser : <https://votre-serveur/OverHAL.php>

Référencer des publications dans HAL

A partir d'un fichier source CSV ou XML, OverHAL repère les publications absentes de HAL et propose selon le cas un export au format TEI ou bibtex, adapté au logiciel d'import [X2HAL](#) (Inria).

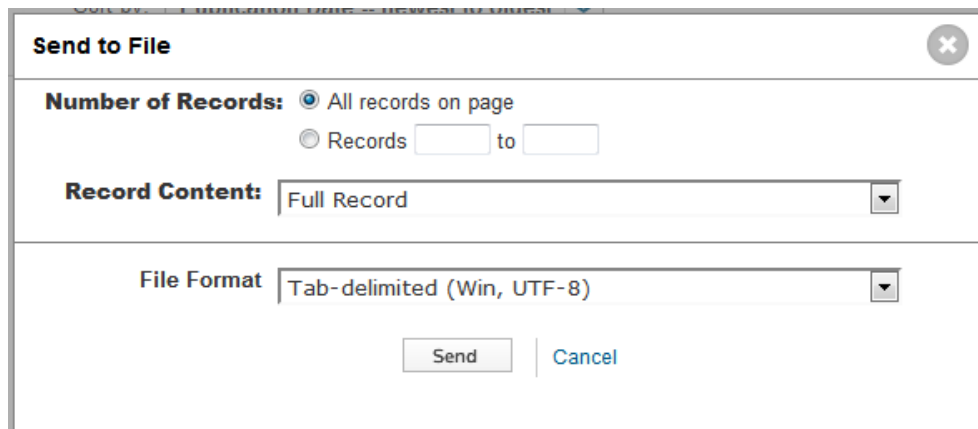
La 1ère chose à faire est de se construire une liste de requêtes pour chaque ressource.

A partir d'une liste de résultats, on fait un export CSV ou XML.

1. Créez un fichier CSV ou XML à partir du résultat d'une requête :

- **Attention : la taille du fichier doit être inférieure à 500 Ko !**
- WoS : export CSV (Save to other file formats / Full Record / Tab-delimited Win, UTF-8)
- Scopus : export CSV / Specified field to be exported (tout cocher)
- Pubmed : export CSV ou XML (Send to file / XML)
- Zotero : export CSV

Créez un fichier CSV d'une liste de résultats WoS (Web of Science) :



Send to File

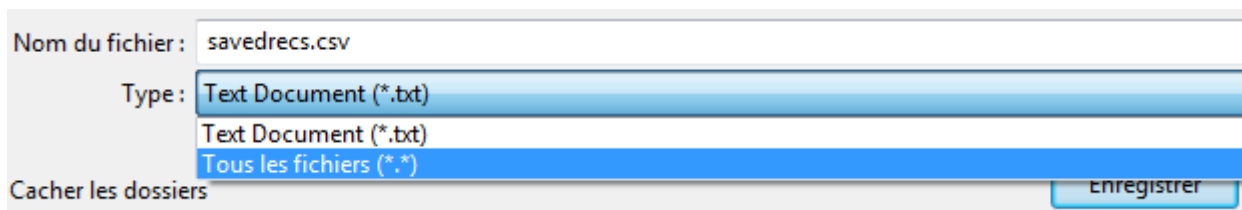
Number of Records: All records on page
 Records to

Record Content: Full Record

File Format: Tab-delimited (Win, UTF-8)

Send | Cancel

Attention : lors de l'enregistrement Windows, le fichier WoS est par défaut en format TXT. Il faut changer l'extension en CSV et choisir comme type « Tous les fichiers » :



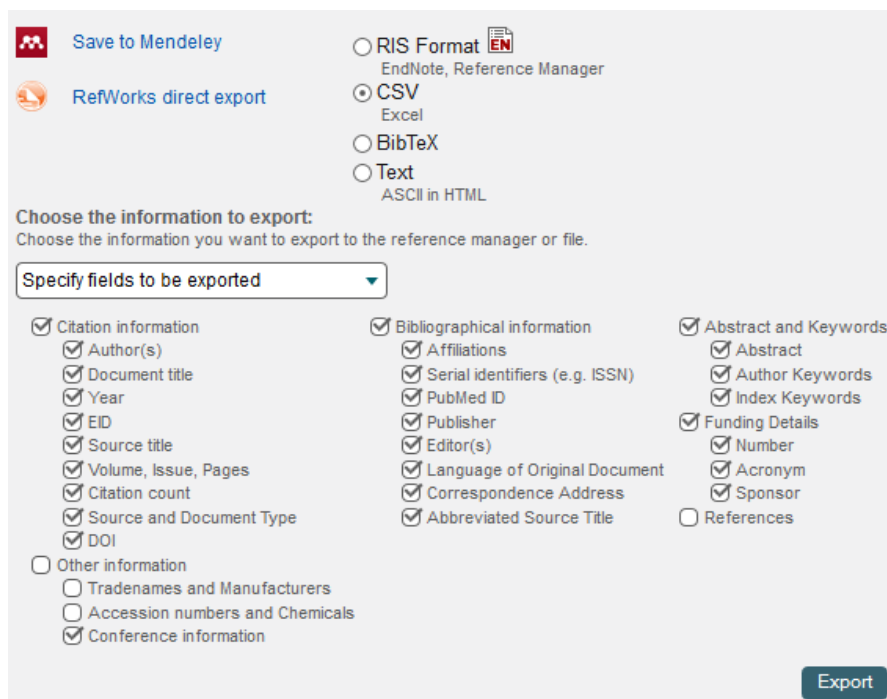
Nom du fichier : savedrecs.csv

Type : Text Document (*.txt)
Text Document (*.txt)
Tous les fichiers (*.*)

Cacher les dossiers

Enregistrer

Créez un fichier CSV d'une liste de résultats Scopus :



Save to Mendeley

RefWorks direct export

RIS Format
EndNote, Reference Manager

CSV
Excel

BibTeX

Text
ASCII in HTML

Choose the information to export:
Choose the information you want to export to the reference manager or file.

Specify fields to be exported

<input checked="" type="checkbox"/> Citation information	<input checked="" type="checkbox"/> Bibliographical information	<input checked="" type="checkbox"/> Abstract and Keywords
<input checked="" type="checkbox"/> Author(s)	<input checked="" type="checkbox"/> Affiliations	<input checked="" type="checkbox"/> Abstract
<input checked="" type="checkbox"/> Document title	<input checked="" type="checkbox"/> Serial identifiers (e.g. ISSN)	<input checked="" type="checkbox"/> Author Keywords
<input checked="" type="checkbox"/> Year	<input checked="" type="checkbox"/> PubMed ID	<input checked="" type="checkbox"/> Index Keywords
<input checked="" type="checkbox"/> EID	<input checked="" type="checkbox"/> Publisher	<input checked="" type="checkbox"/> Funding Details
<input checked="" type="checkbox"/> Source title	<input checked="" type="checkbox"/> Editor(s)	<input checked="" type="checkbox"/> Number
<input checked="" type="checkbox"/> Volume, Issue, Pages	<input checked="" type="checkbox"/> Language of Original Document	<input checked="" type="checkbox"/> Acronym
<input checked="" type="checkbox"/> Citation count	<input checked="" type="checkbox"/> Correspondence Address	<input checked="" type="checkbox"/> Sponsor
<input checked="" type="checkbox"/> Source and Document Type	<input checked="" type="checkbox"/> Abbreviated Source Title	<input type="checkbox"/> References
<input checked="" type="checkbox"/> DOI		
<input type="checkbox"/> Other information		
<input type="checkbox"/> Tradenames and Manufacturers		
<input type="checkbox"/> Accession numbers and Chemicals		
<input checked="" type="checkbox"/> Conference information		

Export

NB : les versions html utilisables dans OverHAL correspondent à des alertes mail (WoS, Pubmed) converties en html. Elles ne sont plus actuellement utilisées et ne sont proposées qu'à titre purement expérimental.

2. Chargez le fichier dans OverHAL

Envoyez les fichiers résultat (500 Ko maximum, voir ci-dessus le "mode d'emploi") :

Web of Science (CSV) : irset-wos.csv

3. Construisez votre requête :

Code collection HAL (qu'est-ce que c'est ?) :

Années de publication : à

- inclure les articles "A paraître"
- requête uniquement sur le texte intégral (dépôt HAL ou lien arxiv ou lien Pubmed Central)
- désactiver les recherches Sherpa/RoMEO

Entrez le nom de la collection HAL ou du portail HAL.
Sélectionnez l'année de publication.

Il est conseillé de faire des requêtes sur une collection ou un corpus HAL contenant moins de 10000 dépôts. Pour une collection plus large, limitez la requête à une année en ajoutant le critère année (exemple : collCode_s:UNIV-RENNES1&fq=producedDateY_i:"2019").

Vous pouvez également modifier la requête par défaut en vous aidant de l'API de HAL :





Requête libre (consultez l'API de HAL)

Filtres disponibles (bouton à cocher) :

- Requête uniquement sur le texte intégral (dépôt HAL ou lien arxiv ou lien Pubmed Central)
- Désactiver les recherches SR (Sherpa-Romeo) : option désactivée par défaut depuis la promulgation de la Loi Numérique qui autorise le dépôt de tous les articles.

URL Joker : option pour récupérer le texte intégral d'une publication : url + DOI

4. Résultats des publications non trouvées dans HAL

Références	DOI	Source	Joker
1 Santos, S; Eekhout, I; Voerman, E; Gaillard, R; Barros, H; Charles, MA; Chatzi, L; Chevrier, C; Chrousos, GP; Corpeleijn, E; Costet, N; <i>et al.</i> (2018) Gestational weight gain charts for different body mass index groups for women in Europe, North America, and Oceania - BMC MEDICINE - doi: https://doi.org/10.1186/s12916-018-1189-1			
2 Lindsey, A; Mozeiko, J; Krueger, F; Grafman, J; Coelho, C (2018) Changes in discourse structure over time following traumatic brain injury - NEUROPSYCHOLOGIA - doi: https://doi.org/10.1016/j.neuropsychologia.2018.08.030			

[Exporter les résultats WoS pour Bib2HAL](#) - [Exporter les résultats WoS au format TEI](#)

Suivant les options cochées, votre tableau de résultats peut inclure les colonnes suivantes :

- Références (avec lien Google Scholar sous le titre de l'article)
- DOI
- Source : lien vers la notice de la base de données, extrait du CSV
- Joker : lien vers un site web de son choix

Informations optionnelles :

- Pprint : politique de la revue dans Sherpa-Romeo (interrogation API) : passez la souris sur l'icône pour afficher le détail.
- PDF : dépôt autorisé / interdit
- DOAJ : si référencé (info Sherpa-Romeo)

5. Exportez en TEI ou bibTex

OverHAL génère un fichier TEI zippé et/ou bibtex adapté à l'application [X2HAL](#).

Faites un clic droit sur le lien « Exportez les résultats WoS au format TEI »

6. TEI ou BibTex ?

TEI est préférable si vous avez pu récupérer les affiliations auteurs dans WoS, Scopus ou Pubmed. Le contrôle des affiliations dans X2HAL en sera facilité.

7. Options supplémentaires

- générer un bibtex de publications en français, à partir d'un fichier Zotero, avec éventuellement,
- . ces mots-clés (séparés par ";"):
 - . cet auteur ("Mortier, Renaud", par exemple):

Cette option permet de générer un bibtex avec par défaut langue = français + audience = national.

Elle a été conçue pour les publications de laboratoires qui ne publient qu'en français, et pour lesquels un référencement par auteur est plus approprié (ex : Doctrinal Plus).

Récupération du texte intégral

Objectif : à partir d'un fichier CSV, générer l'envoi de mails à un ou plusieurs auteurs d'une publication afin d'obtenir :

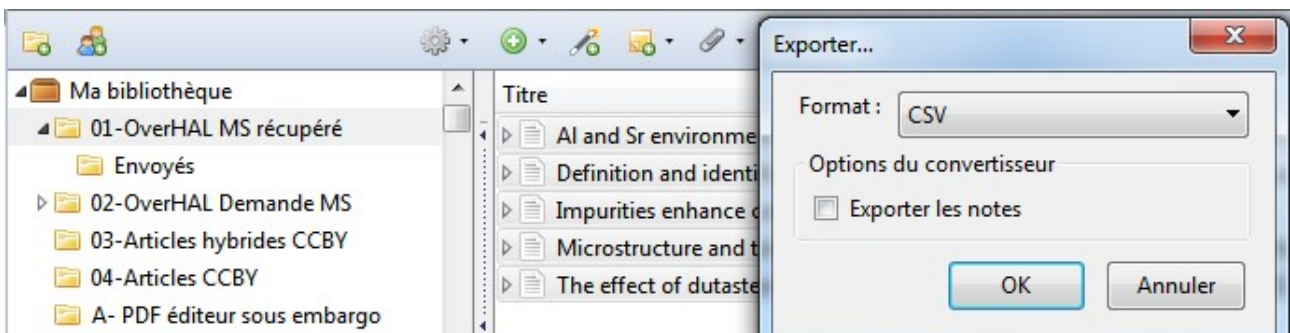
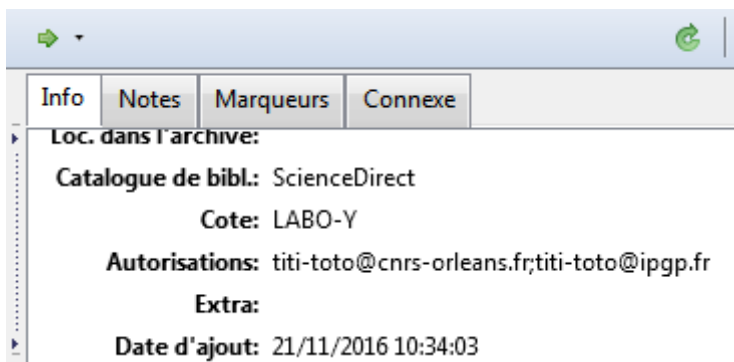
- soit le post-print (option M)
- soit l'autorisation de déposer le post-print (option P).

A Rennes 1, cette fonctionnalité est utilisée pour tous les articles « in press » repérés dans notre veille sur les plateformes éditeurs (Elsevier, Wiley, Springer, etc.). Elle fonctionne avec des CSV de Zotero et Scopus (WoS ne référence pas les articles in press).

1. Créez le fichier CSV

Scopus : en recherche avancée, ajouter le critère « AND DOCTYPE (ip) » à votre requête pour limiter l'affichage aux articles in press. Export CSV / Specified field to be exported (tout cocher)

Zotero : ajouter manuellement pour chaque article l'adresse email de l'auteur correspondant (champ Autorisation ; adresses séparées par des points-virgules ; 3 adresses maximum) et l'acronyme du labo défini lors de l'installation dans les fichiers « OverHAL_mails_xxx.php » (champ Cote). Pour faciliter cette opération et éviter les erreurs, nous conseillons de classer les articles dans 2 répertoires distincts de la bibliothèque Zotero : Manuscrits récupérés / Demandes de manuscrits.



NB : A Rennes 1, nous stockons les manuscrits post-prints dans Zotero (Préférences > Joindre les PDF). Si l'auteur répond positivement, il suffit de rechercher l'article dans le répertoire Zotero « Envoyés » à l'aide du DOI et de récupérer le PDF.

2. Chargez le fichier dans OverHAL

Zotero (CSV) : ur1-ms.csv

3. Construisez votre requête :

Attention : la fonctionnalité mailing n'est disponible que si vous avez déclaré une ou plusieurs adresses IP dans les fichiers « OverHAL.php » et « OverHAL_mails_envoyes_listing.php » (voir procédure d'installation). Il est nécessaire également de renseigner l'adresse mail de l'expéditeur des mails et de la personne en copie des mails envoyés aux auteurs (en principe la même adresse) dans le fichier « OverHAL_mailto.php ».

Code collection HAL (qu'est-ce que c'est ?) :

Années de publication : à

- inclure les articles "A paraître"
- requête uniquement sur le texte intégral (dépôt HAL ou lien arxiv ou lien Pubmed Central)
- désactiver les recherches Sherpa/RoMEO
- activer les procédures d'envoi de mails (demande du manuscrit à l'auteur = M)
- activer les procédures d'envoi de mails (autorisation de dépôt du post-print = P)

Entrez le nom du portail HAL.
Sélectionnez l'année de publication.

Cochez l'option :
« Requête uniquement sur le texte intégral (dépôt HAL ou lien arxiv ou lien Pubmed Central) »










S'il s'agit de publications pour lesquelles **vous avez récupéré le manuscrit**, cochez :
« Activer les procédures d'envoi de mails (demande du manuscrit à l'auteur = **P**) »

S'il s'agit de publications pour lesquelles **vous souhaitez obtenir le manuscrit**, cochez :
« Activer les procédures d'envoi de mails (demande du manuscrit à l'auteur = **M**) »

4. Résultats

Dans la colonne Mails, **cliquez sur le bouton M ou P** pour envoyer les mails pré-rédigés aux auteurs correspondants des publications. Le mail sera envoyé aux auteurs, avec votre adresse en copie.

Si un bouton apparaît en vert : l'adresse mail est absente du fichier CSV. Il faut la renseigner et recharger le fichier dans OverHAL. Si « **OK** » s'affiche dans la colonne : un mail a déjà été envoyé à l'auteur pour ce DOI. Le bouton M ou P ne sera pas actif pour les auteurs enregistrés dans le fichier « auteurs à exclure » (voir procédure d'installation).

	DOI	Source	Joker	Mails
J.-C., Leclercq C., Mabo P., Verhoye J.-P., Martins R.P. (2016) Effectiveness of Shock Due To Intractable Arrhythmic Storm - <i>Critical Care Medicine</i> - doi:				
eber B., Collange O., Dewitte A., Martin-Lefevre L., Jabaudon M., Kerforne T., au S., Lecron J.-C., Robert R., and The ABDOMIX group (2016) Modulation by Related to Severe Peritonitis - <i>Shock</i> - doi: https://doi.org/10.1097				
retagne J.-F., Brochard C. (2016) Prevalence and characteristics of acid gastro- and Liver Disease - doi: https://doi.org/10.1016/j.dld.2016.06.030				
A weird polyp, 8 years after the Whipple procedure - <i>Surgery (United States)</i> -				

OverHAL envoie par défaut un mail en anglais, sauf si l'adresse mail de l'auteur se termine par « .fr ». S'il y a plusieurs adresses, avec au moins une adresse sans .fr, le mail sera en anglais.

Liste des auteurs contactés

Les envois de mails sont enregistrés dans une liste auto-générée. Quand un auteur a déjà été contacté dans les 6 derniers mois, le contenu du mail envoyé diffère légèrement : « Je me permets de vous écrire à nouveau au sujet d'un article dont vous êtes co-auteur... ». Les données de la liste sont sauvegardées 6 mois pour ne pas ralentir l'exécution du programme.

On peut consulter, compléter et amender cette liste dans son navigateur web (voir procédure d'installation) : https://votre-serveur/OverHAL/OverHAL_mails_envoyes_listing.php

Personnalisation des emails

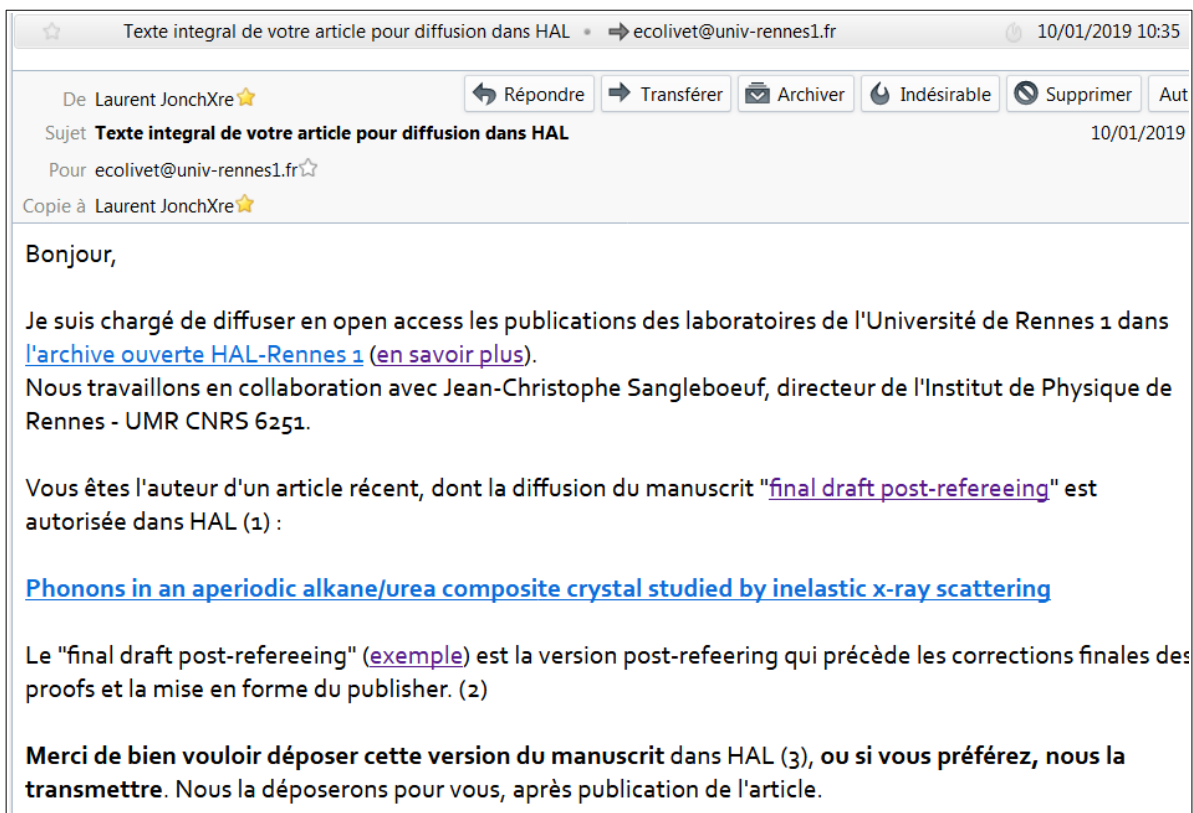
Option : vous pouvez personnaliser votre mail en y insérant une phrase-type qui sert en quelque sorte de recommandation (voir procédure d'installation) :

« Nous travaillons en collaboration avec Jacques Dupont, directeur de l'Institut Y »

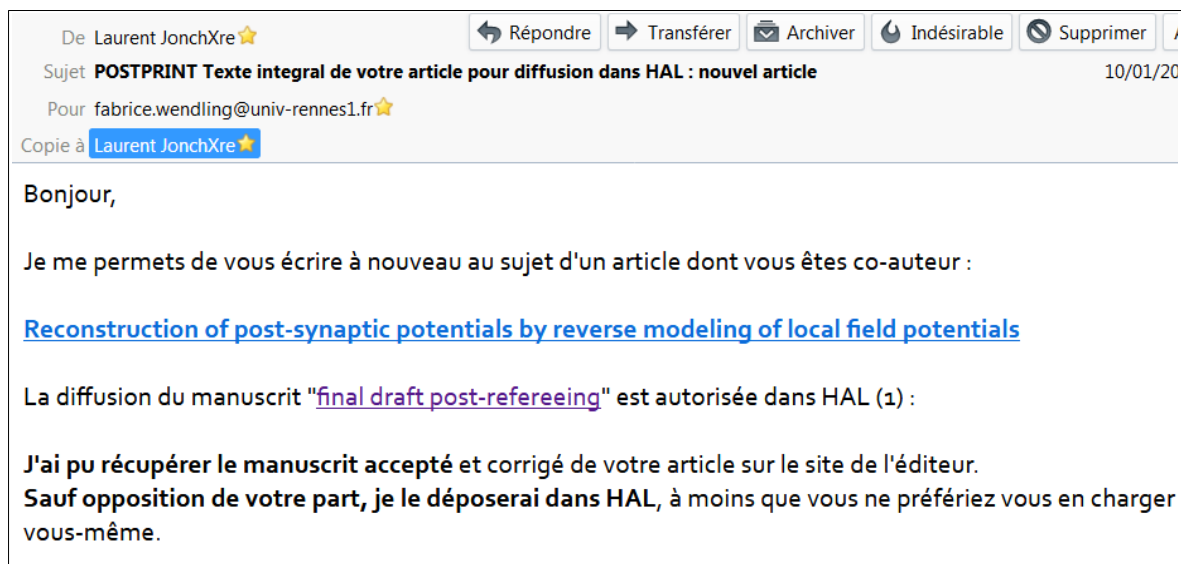
5. Réception des messages

L'envoi des mails est activé depuis l'application. L'expéditeur reçoit une copie des envois s'il a correctement enregistré son adresse mail dans le fichier « OverHAL_mailto.php » lors de l'installation.

Exemple d'un message de type **M** (demande d'envoi du manuscrit) :



Exemple d'un message de type **P** (demande d'autorisation de dépôt du manuscrit) :



Nous avons privilégié des messages très synthétiques, pas trop longs, sur le modèle : qui je suis, la légitimité de ma démarche (je travaille avec untel, donc ce n'est pas du spam), ce que je veux, sous quelle forme, en laissant pour la fin l'argumentaire de base (visibilité, etc.).

6. Gestion des messages

Procédure à Rennes 1 :

- **Réponse positive de l'auteur**, approuvant le dépôt ou communiquant son manuscrit : le message est archivé dans un répertoire de messagerie : « à déposer ». Le message initial est supprimé ou archivé dans un dossier « traité » pour éviter une relance.
- **Pas de réponse au bout de n jours** : si message de type P, nous déplaçons le message initial dans le répertoire de messagerie : « à déposer ». Si message de type M, nous envoyons un message de relance à partir d'un modèle (via [Quicktext](#) pour Thunderbird). Il n'y a pas de suivi sur les relances.
- **Réponse négative de l'auteur** : si nous n'avons pas réussi à convaincre l'auteur de revenir sur son refus, le message initial est supprimé ou archivé dans un dossier « refus » pour éviter une relance. L'adresse mail de cet auteur est éventuellement ajoutée dans le fichier PHP « mails à exclure » pour éviter de le contacter à nouveau.

Outre la récupération de manuscrits, cette procédure a pour intérêt de pouvoir engager des échanges avec les auteurs sur des questions souvent récurrentes (a-t-on le droit ? Quelle version ? Etc.). Là encore, le module [Quicktext](#) pour Thunderbird se révèle diablement utile pour apporter des réponses rapides, sans devoir chaque fois réécrire la même argumentation.

Pour plus d'efficacité et de meilleurs retours, nous conseillons aux utilisateurs d'informer au préalable les chercheurs de votre établissement de votre démarche, afin que vos messages ne soient pas considérés comme du spam. A Rennes 1, nous avons profité de réunions open access dans les laboratoires pour faire cette information, et constaté un effet positif sur les contributions.

7. Quels auteurs contacter ?

A Rennes 1, pour les messages de type P (manuscrit récupéré sur le site de l'éditeur ou par un autre moyen) : nous contactons généralement le ou les auteurs du laboratoire affilié à notre université.

Pour les messages de type M (demande d'envoi du manuscrit), en revanche, nous préférons contacter directement l'auteur correspondant, censé détenir la version finale acceptée, voire corrigée, de la publication.

ANNEXE 1 : Problèmes techniques référencés

- **Le fichier zip (TEI) est vide** : essayez d'ouvrir le fichier avec l'outil de compression 7z (Z-zip) si l'outil Windows ne fonctionne pas.
- **Le fichier CSV est mal structuré** : cela peut se produire notamment pour un export Scopus : certaines données sont décalées. Il faut donc corriger le CSV (couper-coller les données dans les bonnes cellules).
- **Le fichier zip (TEI) contient un fichier xml sans numéro de référence** « OverHAL_scopus_.xml » : ouvrez le zip et supprimez le fichier,

ANNEXE 2 : Questions techniques

- **Transformation du CSV en XML-TEI conforme** : comment est faite la transformation de l'un à l'autre ? (question posée lors d'un atelier CasuHAL) : on réutilise simplement les métadonnées du CSV (titre, DOI, pagination, etc) pour les organiser d'une manière différente dans le XML. On part d'une structure XML vide et valide XSD que le script remplit au-fur-et-à-mesure qu'il récupère les informations du CSV. L'étape intermédiaire consiste en la construction d'une chaîne en PHP, complétée par les variables du CSV et finalement enregistrée comme fichier XML.

Voilà le début de la construction de la chaîne :

```
$chaîne = "";
$chaîne .= '<?xml version="1.0" encoding="UTF-8"?>'. "\r\n";
$chaîne .= '<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:schemaLocation="http://www.tei-c.org/ns/1.0 http://api.archives-ouvertes.fr/documents/aofr-sword.xsd">'. "\r\n";
$chaîne .= ' <text>'. "\r\n".
' <body>'. "\r\n".
' <listBibl>'. "\r\n".
' <biblFull>'. "\r\n".
' <titleStmt>'. "\r\n";
//funder
$chaîne .= ' <funder>'.supprAmp($papers[$key][$key2]['FU']). '</funder>'. "\r\n";
```

Liens utiles :

- page du schéma xsd complet de HAL : <https://hal.archives-ouvertes.fr/documents/aofr.xsd>
- TEI complet de HAL (avec doc intégrée) : <https://api.archives-ouvertes.fr/documents/all.xml>
- pour afficher le TEI d'une notice HAL, ajouter /TEI à la fin de l'url : <https://hal-univ-rennes1.archives-ouvertes.fr/hal-01937242/TEI>
- quelques exemples de TEI par type de document (point 3.4) : <https://api.archives-ouvertes.fr/docs/sword>

A noter qu'OverHAL interroge via SWORD le référentiel des structures pour établir une correspondance, quand c'est possible, entre l'affiliation WoS / Scopus / Pubmed et la structure référencée dans HAL.