



HAL
open science

SPIP dans la recherche : la gestion avec RSS des listes de publications déposées au Centre pour la Communication Scientifique Directe (CCSD/HAL)

Minh Ha-Duong

► **To cite this version:**

Minh Ha-Duong. SPIP dans la recherche : la gestion avec RSS des listes de publications déposées au Centre pour la Communication Scientifique Directe (CCSD/HAL). 2006. halshs-00008414

HAL Id: halshs-00008414

<https://shs.hal.science/halshs-00008414>

Submitted on 30 Jan 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SPIP dans la recherche : la gestion avec RSS des listes de publications déposées au Centre pour la Communication Scientifique Directe (CCSD/HAL)

Minh Ha-Duong, CIRED, CNRS

26/01/06, 13 pages, 3808 mots, 24679 signes

Résumé

Les listes de publications scientifiques sont un élément clé des sites de laboratoires de recherche. Elles sont aussi essentielles aux pages personnelles des chercheurs. Gérer ces listes de façon systématique implique de connecter d'une façon ou d'une autre un serveur web avec une base de donnée des productions scientifiques. Cette note expose comment connecter l'archive ouverte HAL et le système de gestion de contenu SPIP par le protocole RSS.

1. Introduction: le modèle *n*-tiers

En bref, ce texte explique comment obtenir l'illustration 1 à partir de l'illustration 2. Le but est d'intégrer les listes de publication sur un site de laboratoire sous SPIP. L'illustration 1 montre le résultat obtenu sur le site de test du CIRED. Conformément aux principes discutés dans Ha-Duong et Mardon¹, on s'impose d'utiliser une version standard de SPIP. Une version postérieure à Janvier 2006 est nécessaire.

Le prérequis le plus lourd est que la liste des publications soit hébergée sur le système HAL² du Centre pour la Communication Scientifique Directe (CCSD). L'illustration 2 ci dessous montre une vue des notices et documents stockées sur l'archive ouverte HAL. Il s'agit du début de la même liste de publications que l'illustration 1, mais vue sur le site de HAL. Une différence notable avec la précédente est que HAL classe en tête les derniers articles entrés en base, alors que sur l'illustration 1 les publications sont triés par nature et classées par année.

Techniquement, le principe est de récupérer la liste de publication auprès de HAL sous la forme d'un fichier RSS. Le site web du laboratoire, motorisé par le système de gestion de contenu SPIP, se contente de la mettre en

1 Ha-Duong et Mardon (2004) Sur la communication par le web d'un laboratoire de recherche. En ligne à <http://www.centre-cired.fr/forum/article168.html>

2 ou HAL-SHS, c'est la même base de données.

Mes publis sur HAL-SHS

Articles dans des revues scientifiques (11)

- Risk aversion, intergenerational equity and climate change. (2004) Ha-Duong, Treich. Full reference...
- A strategy for bounding attributable risk: a lung cancer example. (2004) Ha-Duong, Casman, Morgan. Full reference...
- Carbon storage: the economic efficiency of storing CO2 in leaky reservoirs (2003) Ha-Duong, Keith. Full reference...
- A pro-active stratospheric ozone protection scenario (2003) Ha-Duong, Mégie, Hauglustaine. Full reference...
- Transparency and control in engineering integrated assessment models. (2001) Ha-Duong. Full reference...
- INASUD project findings on integrated assessment of climate policies (2001) Hourcade, Ha-Duong, Grubler, Tol. Full reference...
- L'initiative européenne "Suivi global pour l'environnement et la sécurité" et le protocole de Kyoto (2000) Bégni, Ha-Duong, Lacaux, Lambin, Le Treut, Morel, Raschke, Saugier, Vellinga, Dekker, Van Der Woerd, Gupta, Zerbini. Full reference...
- Perspectives sur le Changement climatique (2000) Ha-Duong. Full reference...
- Dynamic consistency problems behind the Kyoto protocol (1998) Ha-Duong, Hourcade, Lecocq. Full reference...
- Quasi-option value and climate policy choices (1998) Ha-Duong. Full reference...
- Influence of socioeconomic inertia and uncertainty on optimal CO2-emission abatement (1997) Ha-Duong, Grubb, Hourcade. Full reference...

Articles dans des ouvrages scientifiques (3)

- An abrupt stochastic damage function to analyse climate policy benefits (2004) Ha-Duong, Dumas. Full reference...
- Policy framework and systems management of global climate change (2002) Ha-Duong, Hourcade. Full reference...
- Necessities and problems of coupling climate and socioeconomic models for integrated assessments studies

Illustration 1.: Présentation d'une liste de publications



The screenshot shows the HAL search interface. At the top, there are logos for YPST, HAL, and CCSD. Below the logos is a navigation bar with tabs for Accueil, Déposer, Consulter, Rechercher, and Services. The 'Rechercher' tab is active. Below the navigation bar, there is a search form with the following fields:

- auteur (nom): Ha-Duong
- est exactement: [dropdown]
- nom du laboratoire: [dropdown]
- est exactement: [dropdown]
- titre: [dropdown]
- contient: [dropdown]
- identifiant arXiv: [dropdown]
- est exactement: [dropdown]

Buttons for 'Rechercher' and 'Annuler' are present. Below the search form, there is a section for 'Trier les résultats suivant' with a dropdown for 'date publication, écriture' and a 'desc.' dropdown. A checkbox for 'voir également les notices bibliographiques (liste de publication)' is also visible. The search results show 22 documents found. The first three results are:

- Review of risk and uncertainty concepts for climate change assessments including human dimensions**
Ha-Duong M.
[halshs-00006089 – version 1] (21/01/2006)
- Modèles de précaution en économie: introduction aux probabilités imprécises**
Ha-Duong M.
[halshs-00007520 – version 1] (23/12/2005)
- Climate strategy with CO2 capture from the air**
Keith D., Ha-Duong M., Stolaroff J.
[halshs-00003926 – version 1] (04/05/2005)

Illustration 2.: Les données sources dans HAL.

forme pour les présenter aux internautes de façon intégrée au site du laboratoire.

C'est une architecture *n*-tiers : pour que l'utilisateur final voie la liste de publication, l'information circule de façon coordonnée entre *n* ordinateurs (*n* indiquant un nombre entier indéfini plus grand que deux-trois, plusieurs). Il y a l'ordinateur devant lequel l'internaute est assis. Le serveur web du laboratoire, qui exécute SPIP et peut (c'est courant) utiliser une base de donnée MySQL sur une autre machine du laboratoire. La base du CCSD est hébergée sur une ou des machines dans un centre de calcul Lyonnais.

2. Implémentation

La mise en oeuvre comporte trois volets. Un, il faut que le webmestre principal configure le SPIP et personnalise ses squelettes pour afficher de belles listes de publis. Deux, il faut configurer une requête pour chaque liste qu'on veut afficher. Trois, et c'est le plus important, il faut initialiser la base de publis dans HAL et la tenir à jour.

2.1 Configurer le site du labo

L'administrateur du site SPIP doit :

- Sur la page « Configuration », « Contenu du site », au niveau du bloc « Référencement de sites et syndication », activer les options « Gérer un annuaire de sites Web », « Utiliser la syndication automatique » et « Ne pas bloquer les liens issus de la syndication ».
- Par défaut, SPIP vérifie les dates de publication dans les flux RSS, et les rejette si elles sont plus anciennes qu'un an. Désactiver cette vérification en mettant la directive `$controler_dates_rss = false` dans le fichier de configuration. Il s'agit du fichier nommé `mes_options.php3` dans le répertoire `spip/ecrire`. Si ce fichier n'existe pas (cas d'une nouvelle installation de SPIP), il faut le créer et il contiendra :

```
<?php
$controler_dates_rss = false;
?>
```

- À mes yeux, le flux envoyé par HAL en ce moment comporte deux défauts cosmétiques. Le premier est que l'identifiant est ajouté à la fin du titre entre crochets, par exemple « La vie des grenouilles en Basse-Picardie [halshs-0000666] ». Le second défaut est que les noms d'auteurs sont formatés « Nom, Prénom », ce qui fait désordre quand on fait une liste d'auteurs séparés par une virgule « Nom1, Prénom1, Nom2, Prénom2, Nom3, Prénom3 ».

Un filtre a été écrit pour de modifier le flux envoyé par HAL en enlevant l'identifiant du titre et les prénoms. Techniquement, c'est un plugin de SPIP qui s'installe au point d'entrée `pre_syndication`.

Cette fonctionnalité n'est disponible qu'à partir de la version 1.9 alpha. En attendant une interface conviviale d'installation des plugins dans SPIP, pour installer ce filtre il faut :

1. Créer un répertoire nommé `plugins` à la racine de SPIP
2. Dans ce répertoire, créer un sous répertoire nommé `RSSdeHALversSPIP`
3. Dans ce sous répertoire, créer un fichier nommé `version.php` contenant le code listé à l'annexe 1
4. Ajouter la ligne `$plugins[] = 'RSSdeHALversSPIP';` à `mes_options.php3`

Le filtre est à adapter si le fichier RSS émis par HAL change, ou si on veut syndiquer un autre dialecte de RSS. Pour optimiser on pourrait déporter la définition de la fonction `cleanHAL()` hors du fichier `version.php`, mais ça devient un peu lourd pour deux lignes de code.

- Enfin il faut personnaliser les squelettes de `site.html` et éventuellement la feuille de style pour présenter de belles listes. L'annexe 2 le code du squelette correspondant à la figure 1. La liste est classée par type de document, puis triée par année de publication.

2.2 Configurer les listes de publication


Dans l'espace privé de SPIP, cliquer sur 'Référencer un nouveau site'. On peut utiliser le référencement automatisé, mais afin de mieux contrôler la présentation il est préférable d'indiquer manuellement le nom du site, son adresse et la description comme dans l'illustration 3 ci dessous. L'URL de la liste de publications est donnée à SPIP comme adresse du fichier « backend ». Voir sur l'illustration le champ associé au bouton « Syndication ». Les paramètres de cette URL décrivent la liste désirée.

En ajustant cette URL , on peut restreindre la liste à un auteur, un labo, une sélection manuelle de publications marquées à cet effet (ce que HAL appelle *un tampon*), aux documents en texte intégral. Les options disponibles sont expliquées sur le le site de HAL, encadré 1.

Les possibilités d'interrogation de HAL par le biais du script `rss.php` sont encore assez sommaires. On note par exemple l'absence de possibilité de limiter la requête sur l'année de publication. Un découpage par année est élémentaire à programmer en SPIP.


2.3 Remplir la base de donnée de publications

C'est le point incontournable et le plus difficile. La documentaliste du laboratoire est au centre de l'action. Au CIRED, nous promovons une approche répartie, c'est à dire que nous demandons aux chercheurs d'archiver eux-même leurs publications. Pour remplir la base, nous avons organisé campagne d'action en faveur de l'auto-archivage.


 Référencer le site :
Mes publis sur HAL-SHS

Nom du site [Obligatoire]

Adresse du site [Obligatoire]


À l'intérieur de la rubrique :

[N'oubliez pas de sélectionner correctement ce champ.]

Description du site

Pas de syndication
 Syndication : ?

Adresse du fichier « backend » pour la syndication :

Illustration 3.: Configuration de la liste des publications

Format rss 2.0 + métadonnées Dublin Core, encodage utf-8
 Sans paramètre, le canal rss vous permet de vous abonner aux nouveautés (nouveau dépôt ou nouvelle version) du jour sur l'archive.
 Les paramètres suivants vous permettent de vous abonner à davantage d'articles :
from *daily* [défaut], *weekly*, *monthly*, *yearly*, *all*. `rss.php?from=monthly` permet de s'abonner aux articles déposés sur l'archive depuis un mois.
format *fulltext* [défaut], *withoutfile* (accès aux dépôts en notices également)
author chaîne de caractères contenant le nom de l'auteur à rechercher (nom ou prénom_nom). `rss.php?author=alain_martin` permet de s'abonner aux articles déposés aujourd'hui dont un des auteurs est *Alain Martin*
lab chaîne de caractères contenant le nom (sigle, nom, affiliation) du laboratoire à rechercher
tampon chaîne de caractères ou tableau contenant les codes des tampons à rechercher
domain chaîne de caractères ou tableau contenant les codes domaines à rechercher

Encadré 1: Syntaxe de l'URL de requête HAL. Source: <http://hal.ccsd.cnrs.fr/>, sur la page d'accueil, survoler avec le pointeur de la souris le RSS sur fond orange à droite dans la barre de menu, le 26/01/2006.

Cette campagne a commencé par une propagande informelle mais franche dans les réunions et les couloirs, et a été concrétisée par une journée spéciale d'action directe « pour augmenter le facteur d'impact du laboratoire ». Il s'agissait de prendre un rendez-vous individuel pour faire archiver un document par le chercheur, à son poste de travail. L'expérience montre qu'en la matière, les promesses sont vaines. Une démarche physique volontaire, directe et personnelle est nécessaire. Les collègues ont généralement toléré de se faire forcer la main, sinon nous avons expliqué la possibilité de faire faire l'archivage par une tierce personne, et extorqué la promesse d'envoyer les textes par email (pour sauver la face, sachant bien que les contrats conclus par contrainte ne sont pas valides).

Pour remplir les archives ouvertes, une des difficultés techniques est qu'une fois qu'on a retrouvé le fichier, il manque souvent le résumé en français et en anglais. Il manque aussi souvent les mots clés, mais je considère que le concept de mot-clé est un relique d'un âge ancien, avant les ordinateurs. Il évoque dans ma mémoire la salle des catalogues de la bibliothèque Sainte-Geneviève, avec ses rangées de meubles à petits tiroirs, suivis des longues attentes pour obtenir le document demandé. C'était sympa, mais maintenant qu'on recherche en texte intégral dans le titre, l'abstract et le texte même du document, on ne se prend plus la tête pour le choix des mots clés.

Le concept de 'guide de bonne conduite' nous semble intéressant pour répondre aux interrogations morales. Dans notre communication aux collègues, nous avons insisté sur les points suivants :

- L'auto-archivage concerne les documents scientifiques (par préférence décroissante: articles soumis ou publiés, communications, mémoires et thèses...)
- Pour les textes publiés, l'indexation de la référence bibliographique se fait au vu du document original.
- Mais on n'archive pas de fac-simile (copie numérisée) d'un article publié, sauf autorisation écrite. D'autant moins qu'il suffit parfois de demander cette autorisation pour l'obtenir.
- On recherche franchement le DOI ou le lien vers la version commerciale de l'article.
- Le format PDF est préférable, le format d'édition (TEX, RTF, DOC...) est aussi acceptable. En pratique, la majorité de nos dépôts sont en PDF, avec quelques textes en HTML.
- Un résumé en français est nécessaire, c'est la loi. Un résumé en anglais ou dans la langue originale du texte est souhaitable.
- L'accord explicite des co-auteurs est préférable à leur accord implicite.

Nous n'avons pas insisté, mais pas caché non plus, la possibilité de déposer seulement une notice et non un texte intégral, ni celle d'archiver avec embargo (1 collègue sur 8 déposants a utilisé cette possibilité).

Un mois après cette campagne, les pratiques évoluent. Le laboratoire est passé de 2 déposants à 8, et de 25 notices à un peu plus de 50. Mais il reste encore plus de la moitié des chercheurs qui n'ont pas encore utilisé ou créé de compte sur HAL. Il est donc nécessaire de garder la pression. Notre objectif est d'indexer 100% de la production scientifique de l'année 2005 et de suivre la production de l'année courante. Concernant l'avant 2005, nous avons lancé une action de rétro-archivage pour mettre en ligne une sélection de textes.

3. Discussion: avantages, inconvénients, alternatives

3.1 Avantages

Il est évidemment rationnel d'avoir une base de donnée unique des publications. On bénéficie d'effets de réseau, puisque les publis avec auteurs multiples n'ont besoin d'être entrés qu'une seule fois. On peut ensuite générer autant de listes qu'on veut, chaque liste correspondant à une requête particulière de la base de donnée. Le système favorise une constitution collective décentralisée de la liste des productions scientifiques du laboratoire. Elles sont plus facilement à jour.

Les avantages d'héberger cette base de donnée primaire dans l'archive institutionnelle et non dans le site du laboratoire sont multiples. Premièrement, c'est en accord avec le schéma directeur informatique du CNRS qui a supprimé PubliCNRS au profit de HAL. Une fois dans HAL, les données peuvent être réutilisées par l'institution de tutelle des chercheurs et des labos dans leurs rapports d'activité.

Deuxièmement, l'archive institutionnelle offre une certification de la date de dépôt (utile pour les questions d'antériorité) plus crédible qu'un site de labo, et un espoir de pérennité supérieure. Elle a l'envergure pour s'intégrer dans les archives ouvertes disciplinaires, ce qui serait une charge bureaucratique inutile pour le webmestre du laboratoire. En économie, HAL est agrégé par WopEc, on bénéficie donc des services de ce réseau comme les statistiques d'accès et les classements.

Troisièmement, cela prévient une dérive possible du site web de labo vers un outil de suivi industriel de la production scientifique. SPIP n'est pas conçu pour. Par contre, il est plus performant que HAL en temps de réponse pour afficher la liste, car il met en cache la liste des publications.

Tout hébergeant les données sur l'archive institutionnelle, conserver la maîtrise de la personnalisation de la présentation et l'intégration au site du laboratoire est avantageux. Les chercheurs sont payés directement avec une liste de publications propre et à jour si ils font l'effort d'archiver ou au moins d'indexer leurs productions sur HAL. A défaut d'une liste intégrale du laboratoire, on peut avoir des résultats même si un seul chercheur utilise le système, ou si on l'utilise pour un seul type de documents. Il est possible de commencer par les thèses, ou par les documents de travail par exemple, et de continuer progressivement.

Enfin l'utilisation d'un protocole standard (RSS est un dialecte de XML) simplifie le travail et garantit la possibilité d'évolutivité. En clair, SPIP et HAL font le gros du travail, le filtre d'adaptation (plugin) fait deux lignes de PHP et peut évoluer en fonction des besoins. Si on ne peut pas encore dire que SPIP et HAL sont les outils standards pour l'usage qui en est fait (respectivement, site de laboratoire dynamique et archive institutionnelle), on peut espérer qu'ils soient en passe de le devenir en France d'ici 2010.

3.2 Risques et inconvénients

L'architecture utilisée suppose la permanence du service de base de donnée du côté de HAL. La permanence à court terme est nécessaire parce que, même si les métadonnées (titre, auteur, résumé...) sont dupliquées du côté de SPIP, le texte intégral en PDF ne l'est pas. Les sont servis à partir de HAL et non du site du labo. Il pourrait être souhaitable que ces documents soient mis en cache localement. Garder une copie locale prémunirait aussi contre une interruption de service du côté de HAL à long terme. Cependant, la meilleure garantie serait plutôt la réplication de la collection par d'autres archives ouvertes.

Le CCSD peut décider de cesser d'offrir le flux RSS. L'archive HAL est une entité en croissance rapide. Ses 40 000 dépôts en font déjà une grosse archive à l'aune des réalisations actuelles, mais s'il capture une fraction significative de la production scientifique française, ce chiffre sera multiplié plusieurs fois. Un tel changement d'échelle ne posera pas que des problèmes techniques, mais aussi institutionnels pouvant remettre en cause la qualité élevée de service et de réactivité obtenue en phase de jeune pousse.

On peut imaginer en particulier que le changement d'échelle appelle une réécriture du code de HAL sur une plateforme Java, ou même une migration vers DSpace ou un autre système de "digital curation" (européen peut être). Dans ce type de turbulences, la continuité des fonctionnalités RSS pourraient passer en arrière plan.

Un risque inhérent au passage de HAL à la masse critique est qu'il devienne un enjeu politique et syndical. L'identité de HAL en tant qu'archive ouverte peut être affectée par son extension à la gestion des notices des productions scientifiques, même sans le texte intégral. Le mot 'flicage' a été entendu dans certains couloirs.

Même à l'intérieur du laboratoire, mettre en base la production d'une organisation a des enjeux humains. L'accroissement de la visibilité des différences de productivité peut être mal vécu. Pour les chercheurs productifs mais peu à l'aise avec les nouvelles technologies, la nécessité de déléguer l'indexation à une tierce personne réactive la relation hiérarchique.

Enfin, ce texte est le produit d'une expérience de terrain qui s'aventure à la pointe des développements logiciels en cours (versions alpha). La méthode exposée n'est pas officiellement sanctionnée par le siège d'une institution. Il

peut être préférable, dans certains contextes bureaucratiques, de s'aligner dans la division régulière plutôt que de participer des francs-tireurs.

3.3 Alternatives

L'architecture proposée pour gérer les listes de publications peut améliorer une situation de non-gestion dans laquelle chaque chercheur se débrouille, et la documentaliste doit faire la liste à la main dans un traitement de texte en catastrophe pour chaque rapport de laboratoire.

L'architecture *n*-tiers nous semble aussi meilleure qu'une gestion centralisée et fermée associée à des logiciels propriétaires souvent utilisés dans le monde de la recherche. En effet, les logiciels de gestion de bibliothèque comme par exemple *Superdoc*, et les logiciels de gestion des références bibliographiques, comme par exemple *EndNotes*, offrent la possibilité d'exporter en HTML une base centralisée de publications. Néanmoins ce n'est pas là la fonction première de ces outils, il leur manque donc la connectivité multidirectionnelle qui caractérise les services web modernes.

HAL offre aussi un accès OAI-PMH, qui est en théorie la même chose que l'accès RSS. En pratique, les notices offertes par HAL par ce protocole ne sont pas aussi complètes que celles offertes par le fichier RSS. De plus, SPIP ne lit pas le OAI en natif (il existe un module SPIP-OAI, mais c'est pour exporter des données et non pour les lire).

HAL offre aussi des fonctions pour exporter une liste publications. Les résultats HTML ne s'intègrent pas aussi bien au site de laboratoire que la solution proposée ici. Le XML produit par HAL est difficilement exploitable car d'une part il ne contient pas toutes les données nécessaires, et d'autre part utiliser XSLT en PHP reste délicat, ce n'est pas intégré à SPIP.

Une solution techniquement intéressante, un peu futuriste peut être, serait que le CCSD rende public le schéma de sa base de données (DTD) et permette de soumettre des requêtes XQUERY. Cela faciliterait grandement l'intégration de HAL avec les autres services web.

Au lieu d'utiliser le plugin de l'annexe 1 pour nettoyer le fichier d'entrée, on aurait pu écrire un proxy PHP. C'est à dire qu'on dit à SPIP que le backend à syndiquer est un script local, charge à ce script d'aller lire la liste des publis sur HAL et la formater en RSS 2.0 idoine.

Une autre alternative aurait été de garder en les champs comme HAL les envoie, et faire le filtrage au niveau du squelette. En effet, SPIP offre la possibilité de définir des filtres personnalisés pour manipuler les textes des balises. La solution d'un plugin a été jugée techniquement meilleure.

Plutôt que de présenter des listes triées côté serveur, on pourrait offrir des listes dynamiques en jouant avec le DOM en javascript pour faire trier la page ou changer la présentation de la liste selon les fantaisies de l'utilisateur. C'est juste faisable avec les navigateurs XML / XSLT / CSS actuels, mais pose des problèmes de compatibilité et cela dépasse le cadre de cet article.

Côté agrégateur, il est clair que SPIP n'est pas l'unique solution non plus. L'idée de passer une requête dans l'URL (dite interface REST) et de récupérer une liste de publication au format RSS peut s'appliquer avec tous les moteurs de site web un peu puissant.

4. Conclusion

Ce texte a présenté une application verticale qui intègre HAL et SPIP au moyen du protocole RSS pour gérer les listes de publications. La méthode présentée est implémentée sur le site de développement du CIRED et sera déployée lors de la mise à jour vers SPIP 1.9. Nous profitons en général du creux estival pour les changements d'infrastructure critique dans le réseau.

La méthode s'inscrit dans une vision de l'infrastructure informatique d'un laboratoire de recherche qui utilise une application spécialisée pour chaque fonction, en exploitant des protocoles d'interconnexion standards.

Ce travail d'intermédiaire a été rendu possible par la bonne volonté des ingénieurs de SPIP et de HAL, je remercie en particulier Fil, Emmanuel, Laurent Capelli et Marin Dacos. Le réseau CIRED est administré par Hoby Ratsihoarana et la documentation par Michèle Sadoun.

Annexe 1: Plugin d'adaptation à SPIP du flux RSS de HAL

```
<?php
/*
  Plugin RSSdeHALversSPIP

  Nettoie le fichier RSS envoyé par HAL pour une belle liste de publication.
*/

$nom = 'RSSdeHALversSPIP';
$version = 0.1;

$GLOBALS['spip_pipeline']['pre_syndication'] .= 'cleanHAL';

function cleanHAL($rss) {
  // Supprime l'identifiant HAL du titre
  $rss = preg_replace('[halshs.*],m', '', $rss);
  // Supprime les prenomms
  $rss = preg_replace('|.*</dc:creator>|m', '</dc:creator>', $rss);
  // On pourrait aussi supprimer les tags inutilisés indiquant la discipline.
  // ...
  return $rss;
}
```

Annexe 2: squelette de présentation de liste de publication

```
<div id="listepublis">

<B_articlesjournal>
  <h2>Articles dans des revues scientifiques (#TOTAL_BOUCLE)</h2>
  <ul>
<BOUCLE_articlesjournal(SYNDIC_ARTICLES){id_syndic}{par date}{inverse}
{tags==ARTJOUR} {doublons}>
  <li><a href="#SOURCE" title="Download !">[(#TITRE*)]</a>
  <span class="detail">
    [(#DATE|annee)]
    #LESAUTEURS.
    <a href="#URL_ARTICLE" class="linktoHAL">Full reference...
    <span>Abstract: [(#DESCRIPTIF|textebrut|entites_html)]</span></a>
  </span>
</li>
</BOUCLE_articlesjournal>
  </ul>
</B_articlesjournal>

<B_articlesouvrage>
  <h2>Articles dans des ouvrages scientifiques (#TOTAL_BOUCLE)</h2>
  <ul>
<BOUCLE_articlesouvrage(SYNDIC_ARTICLES) {id_syndic} {par date}{inverse}
{tags==ARTOUVRAGE}{doublons}>
  <li><a href="#SOURCE" title="Download !">[(#TITRE*)]</a>
  <span class="detail">
    [(#DATE|annee)]
    #LESAUTEURS.
    <a href="#URL_ARTICLE" class="linktoHAL">Full reference...
    <span>Abstract: [(#DESCRIPTIF|textebrut|entites_html)]</span></a>
  </span>
</li>
</BOUCLE_articlesouvrage>
  </ul>
</B_articlesouvrage>

<B_communications>
<h2>Articles dans des actes de colloques (#TOTAL_BOUCLE)</h2>
<ul>
  <BOUCLE_communications(SYNDIC_ARTICLES) {id_syndic} {par
date}{inverse}{tags==ARTCOLLOQUE}{doublons}>
<li><a href="#SOURCE" title="Download !">[(#TITRE*)]</a>
<span class="detail">
  [(#DATE|annee)]
  #LESAUTEURS.
<a href="#URL_ARTICLE" class="linktoHAL">Full reference...
  <span>Abstract: [(#DESCRIPTIF|textebrut|entites_html)]</span></a>
</span>
</li>
</BOUCLE_communications>
</ul>
</B_communications>

<B_articlesautres>
<h2>Articles et communications dans des revues de d&eacute;bat ou d'information
(#TOTAL_BOUCLE)</h2>
<ul>
```

```

        <BOUCLE_articlesautres(SYNDIC_ARTICLES) {id_syndic} {par
date}{inverse}{tags==ARTREVUE|COMMUNICATION}{doublons}>
<li><a href="#SOURCE" title="Download !">[ (#TITRE*) ]</a>
<span class="detail">
        ([ (#DATE|annee) ])
        #LESAUTEURS.
<a href="#URL_ARTICLE" class="linktoHAL">Full reference...
        <span>Abstract: [ (#DESCRIPTIF|textebrut|entites_html) ]</span></a>
</span>
</li>
</BOUCLE_articlesautres>
</ul>
</B_articlesautres>

<B_theses>
<h2>Thèses, mémoires de DEA et d'habilitation (#TOTAL_BOUCLE)</h1>
<ul>
        <BOUCLE_theses(SYNDIC_ARTICLES) {id_syndic} {par
date}{inverse}{tags==THESE}{doublons}>
<li><a href="#SOURCE" title="Download !">[ (#TITRE*) ]</a>
<span class="detail">
        ([ (#DATE|annee) ])
        #LESAUTEURS.
<a href="#URL_ARTICLE" class="linktoHAL">Full reference...
        <span>Abstract: [ (#DESCRIPTIF|textebrut|entites_html) ]</span></a>
</span>
</li>
</BOUCLE_theses>
</ul>
</B_theses>

<B_autres>
<h2>Autres publications (#TOTAL_BOUCLE)</h1>
<ul>
        <BOUCLE_autres(SYNDIC_ARTICLES) {id_syndic} {par date}{inverse}{doublons}>
<li><a href="#SOURCE" title="Download !">[ (#TITRE*) ]</a>
<span class="detail">
        ([ (#DATE|annee) ])
        #LESAUTEURS.
<a href="#URL_ARTICLE" class="linktoHAL">Full reference...
        <span>Abstract: [ (#DESCRIPTIF|textebrut|entites_html) ]</span></a>
</span>
</li>
</BOUCLE_autres>
</ul>
</B_autres>

</div> <!-- listepublis -->

```