

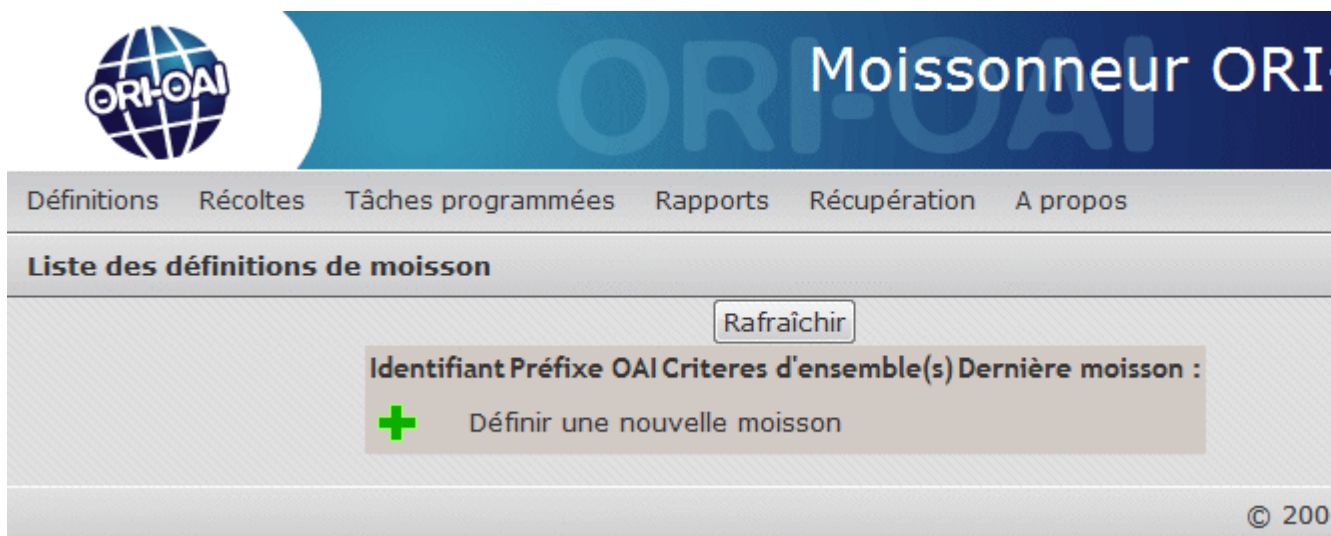
Utiliser ORI-OAI-harvester (Moissonneur)

Utilisation du module ORI-OAI-harvester

Ouvrir un navigateur web et, suivant l'installation effectuée, aller à l'adresse du module.
Si le module a été installé selon la documentation proposée avec un Tomcat par module, cette adresse devrait ressembler à :

```
http://NOM_SERVEUR_ORI:8181/ori-oai-harvester
```

L'utilisation proposée est faite sur une application installée par défaut vide.



L'interface propose 5 onglets :

1. **Définitions** correspond à la page d'accueil, elle permet de déclarer une nouvelle moisson et de gérer les moissons enregistrées
2. **Récoltes** permet de visualiser les récoltes effectuées par le module
3. **Tâches programmées** permet de visualiser les programmations définies pour chaque moisson
4. **Rapports** permet de consulter les rapports des moissons effectuées
5. **Récupération** permet de régénérer l'index
6. **A propos** propose une description du module et des liens vers la documentation

Définitions

L'onglet "définitions" liste les moissons qui ont été définies, et permet de mener les actions suivantes :

- Définir une nouvelle moisson : ouvre la page d'édition d'une définition en mode "création" (voir image plus bas)
- Editer une définition : ouvre la page d'édition d'une définition en mode "édition". Ce mode est plus restrictif que le mode "création" dans le sens où l'on ne peut pas modifier la liste des entrepôts.
- Supprimer une définition : supprime la définition (icone poubelle). Attention, cela ne supprime pas la récolte, c'est-à-dire l'ensemble des fiches moissonnées et indexées : il faut la supprimer AVANT de supprimer la définition, dans l'onglet "Récoltes".
- Lancer une moisson : lance la moisson correspondant à la définition (icone flèche verte).

Définir une nouvelle moisson

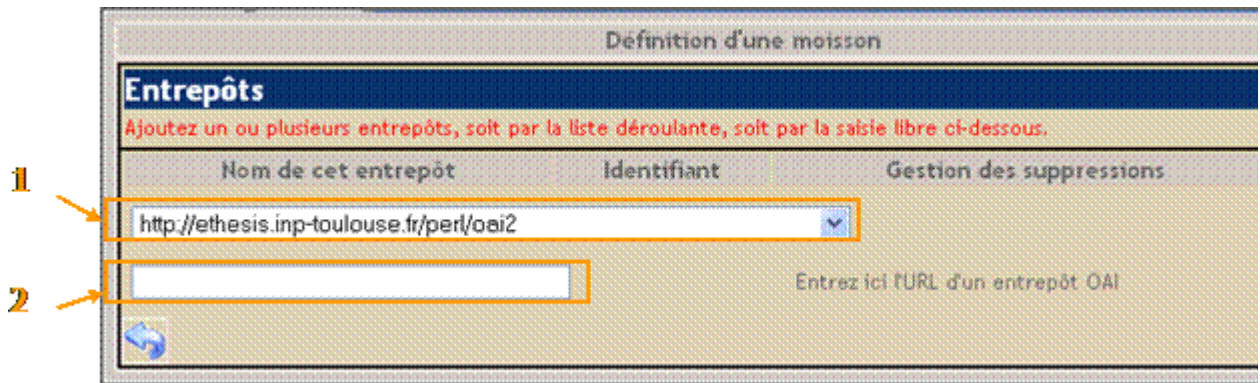


A partir de l'onglet « Définitions », cliquer sur  pour ajouter la définition d'une nouvelle moisson.

1) Déclarer l'entrepôt que l'on souhaite moissonner

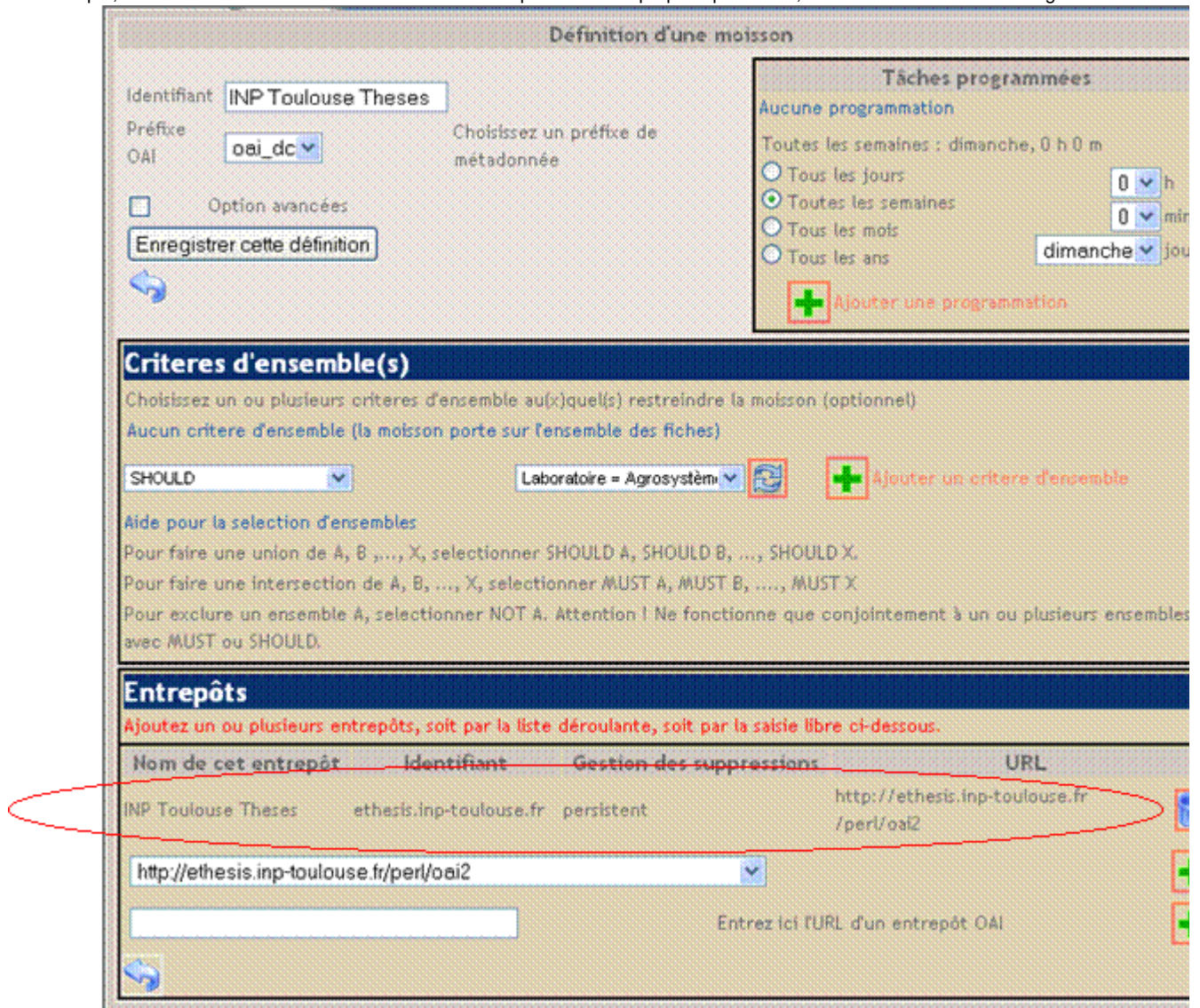
Pour cela, deux possibilités sont proposées :

- [1] sélectionner un entrepôt dans la liste de ceux proposés,
- [2] indiquer l'URL OAI d'un entrepôt non présent dans cette liste (voir effectuer une veille sur les entrepôts à moissonner).



Pour valider le choix effectué, cliquer sur le « + » situé à droite de la ligne choisie. Les informations de l'entrepôt choisi doivent alors s'afficher dans le formulaire.

Par exemple, en sélectionnant dans le menu déroulant l'entrepôt de thèses proposé par l'INPT, vous devriez obtenir l'affichage suivant :



Si les informations ne s'affichent pas correctement (le nom de l'entrepôt a comme valeur « not initialized » par exemple), c'est que l'entrepôt ne peut pas être moissonné (voir partie « Tester un entrepôt avant de le moissonner » au bas de cette page). Il est aussi possible de déclarer une moisson multiple en ajoutant plusieurs entrepôts (voir plus bas).

2) Programmer les moissons dans l'entrepôt

Il existe deux façons de lancer une moisson : le lancement immédiat et le lancement par programmation. Le lancement immédiat s'opère à partir de l'onglet « Définitions », **une fois la définition enregistrée**, à l'aide du bouton vert « Lancer la moisson maintenant » (voir plus bas). En plus de ce mode immédiat et manuel, l'utilisateur peut définir, pour une moisson, une ou plusieurs programmations, en précisant pour

chacune d'elles une périodicité.

Pour cela, il faut utiliser la partie en haut à droite du formulaire.

The screenshot shows a form titled "Tâches programmées". At the top, it says "Aucune programmation". Below that, it displays "Toutes les semaines : dimanche, 0 h 0 m". There are four radio button options: "Tous les jours", "Toutes les semaines" (which is selected), "Tous les mois", and "Tous les ans". To the right of these options are input fields: a dropdown for hours set to "0", a dropdown for minutes set to "0", and a dropdown for days set to "dimanche". At the bottom left, there is a red plus icon in a square followed by the text "Ajouter une programmation".

Sélectionner la périodicité (tous les jours, toutes les semaines, ...), puis détailler votre choix dans les champs proposés. Les informations remplies s'expriment en français au-dessus du formulaire. Pour valider le formulaire, cliquer sur « + Ajouter une programmation ». A la place du message « Aucune programmation », la programmation choisie s'affiche.

This screenshot shows the same form after validation. The text at the top now reads "Toutes les semaines : mardi, 22 h 0 m" and is circled in orange. The "Toutes les semaines" radio button is still selected. The input fields now show "22" for hours, "0" for minutes, and "mardi" for days. The "Ajouter une programmation" button remains at the bottom left.

3) Définir les paramètres de moisson de l'entrepôt

Pour cela, il faut utiliser la partie en haut à gauche et la partie "Critères d'ensemble" du formulaire.

Remplir :


- le champ **identifiant** correspond à l'étiquette que vous souhaitez donner à cette moisson (c'est cette information qui est reprise pour reconnaître une définition de moisson dans les interfaces ; par défaut l'identifiant de l'entrepôt apparaît).
- le champ **préfixe OAI** liste les formats des schémas de métadonnées proposés par l'entrepôt sélectionné ; n'affiche que les préfixes communs à tous les entrepôts.
- Les **Options avancées** permettent de préciser une période de temps (depuis... jusqu'à...) qui va restreindre les fiches moissonnées aux dates de début et de fin indiquées. Elles ne doivent être utilisées que dans le cas où l'on ne veut moissonner que les fiches correspondant à une période précise. Au moment de son prochain lancement, la moisson utilisera ces critères de sélection avancés, qui seront mis à jour **automatiquement** à l'issue de la moisson de la façon suivante : la date de fin sera enlevée et la date de début sera réglée sur la date de fin de la moisson, afin que **les moissons s'effectuent toujours de façon incrémentale**. Afin de satisfaire le choix de restreindre la moisson à une date maximale, aucune itération de la moisson ne doit donc être faite.

Définition d'une

Identifiant

Préfixe OAI Choisisissez un préfixe de métadonnée

Option avancées




- vous pouvez (**optionnel**) restreindre la moisson en sélectionnant des ensembles disponibles dans le ou les entrepôts, dans la partie "**Critères d'ensemble(s)**" (remarque : il est possible de moissonner différents sets d'un entrepôt dans une même définition de moisson) ; si aucun ensemble n'est choisi, tout l'entrepôt est moissonné.

Critères d'ensemble(s)

Choisissez un ou plusieurs critères d'ensemble au(x)quel(s) restreindre la moisson (optionnel)

Aucun critère d'ensemble (la moisson porte sur l'ensemble des fiches)



[Aide pour la sélection d'ensembles](#)

Pour faire une union de A, B, ..., X, sélectionner SHOULD A, SHOULD B, ..., SHOULD X.

Pour faire une intersection de A, B, ..., X, sélectionner MUST A, MUST B, ..., MUST X.

Pour exclure un ensemble A, sélectionner NOT A. Attention ! Ne fonctionne que conjointement à un ou plusieurs ensembles avec MUST ou SHOULD.

Sélectionner le type de sélection et l'ensemble choisi dans les listes déroulantes ; puis ajouter en cliquant sur




4) Valider cette définition de moisson

Définition d'une moisson


Identifiant

Préfixe OAI Choisisissez un préfixe de métadonnée

Option avancées



Tâches programmées

Toutes les semaines : mardi, 22 h 0 m 

Toutes les semaines : mardi, 22 h 0 m

Tous les jours

Toutes les semaines

Tous les mois

Tous les ans







22 h
0 min
mardi jour

Une fois le formulaire complètement rempli, cliquez sur « enregistrer cette définition », vous retournez sur la page d'accueil, de définition des moissons.

Lancer une moisson manuellement

A partir de l'onglet « définitions », sur la ligne correspondant à la moisson que vous souhaitez lancer, cliquez sur le bouton (rond) vert pour lancer la moisson.



Liste des définitions de moisson

Identifiant	Collection	Préfixe OAI	Critères d'ensemble(s)	Dernière moisson :
CEL_Set_SHS	  	oai_dc	SHOULD Humanities and Social Sciences	
UNIT_Set-Informatique	  	lom	SHOULD Informatique	

Définir une nouvelle moisson

Modifier une définition de moisson

A partir de l'onglet « définitions », sur la ligne correspondant à la moisson à modifier, cliquer sur l'icône « écrire ».


Liste des définitions de moisson				
Rafraichir				
Identifiant	Collection	Préfixe OAI	Criteres d'ensemble(s)	Dernière moisson :
CEL_Set_SHS		oai_dc	SHOULD Humanities and Social Sciences	
UNIT_Set-Informatique		lom	SHOULD Informatique	
+ Définir une nouvelle moisson				

Le formulaire de modification s'affiche. Les éléments de définition de la moisson autres que le ou les entrepôts moissonnés et l'identifiant initialement donné sont modifiables, ainsi que la programmation de la moisson.

Définition d'une moisson

Identifiant: CEL_Set_SHS Non modifiable en mode édition
 Préfixe OAI: Choisissez un préfixe de métadonnée
 Option avancées

Tâches programmées

Tous les mois : jour 5, 0 h 10 m 

Toutes les semaines : dimanche, 0 h 0 m


Tous les jours
 Toutes les semaines
 Tous les mois
 Tous les ans


h
 min
 jour

+ Ajouter une programmation

Criteres d'ensemble(s)

Choisissez un ou plusieurs criteres d'ensemble au(x)quel(s) restreindre la moisson (optionnel)

SHOULD Humanities and Social Sciences 



+ Ajouter un critere d'ensemble

Aide pour la selection d'ensembles

Pour faire une union de A, B, ..., X, selectionner SHOULD A, SHOULD B, ..., SHOULD X.

Pour faire une intersection de A, B, ..., X, selectionner MUST A, MUST B, ..., MUST X



Pour exclure un ensemble A, selectionner NOT A. Attention ! Ne fonctionne que conjointement à un ou plusieurs ensembles avec MUST ou SHOULD.

Entrepôts

Nom de cet entrepôt	Identifiant	Gestion des suppressions	URL
CEL - @HAL	ce.archives-ouvertes.fr	no	http://ce.archives-ouvertes.fr/osi/osi.php

Supprimer une définition de moisson

A partir de l'onglet « définitions », sur la ligne correspondant à la moisson à supprimer, cliquer sur l'icône « poubelle » pour supprimer la moisson.

Liste des définitions de moisson				
Rafraichir				
Identifiant	Collection	Préfixe OAI	Criteres d'ensemble(s)	Dernière moisson :
CEL_Set_SHS		oai_dc	SHOULD Humanities and Social Sciences	
UNIT_Set-Informatique		lom	SHOULD Informatique	
+ Définir une nouvelle moisson				

Remarque : cela ne supprime pas les fiches déjà moissonnées sur les entrepôts définis dans cette moisson. Si vous souhaitez les supprimer

il faut le faire également.

Définir une moisson multiple

Il est possible de définir plusieurs entrepôts à moissonner simultanément. Dans ce cas, les éléments de définition sont communs à l'ensemble des entrepôts choisis. Il n'est donc pas possible par exemple de choisir un set particulier ou un format de métadonnées que seul l'un des entrepôts propose.



Récoltes

L'onglet "Récoltes" permet de visualiser le nombre de documents qui ont été récoltés et qui sont stockés dans la base de données.

Consulter les récoltes

Cliquez sur l'onglet « Récoltes », la liste des moissons définies s'affiche, avec pour chacune la date de la dernière moisson effectuée et le nombre global de documents moissonnés.

Cliquer sur le bouton « Rafraîchir » pour mettre à jour les informations affichées.


Liste des récoltes		
Identifiant	Entrepôts	
CEL_Set_SHS 	Dernière moisson : 2009-03-10	Nombre de documents
	préfixe oai_dc	
	Critères d'ensemble(s) SHS	
	cel.archives-ouvertes.fr	11
Total : 11		
UNIT_Set-Informatique 	Dernière moisson :	Nombre de documents
	préfixe lom	
	Critères d'ensemble(s) search_unit_taxonomie_regexp:13	
	www.unit.eu	0
Total : 0		

Une moisson peut durer d'une à plusieurs minutes.

Pour les nouvelles moissons, appuyer sur rafraîchir à intervalle régulier. La moisson est terminée lorsque le compteur « Nombre de documents » ne s'incrémente plus et que la date du jour apparaît dans le champ « Dernière moisson ».

Pour visualiser l'affichage des fiches moissonnées, consulter le moteur de recherche.

Supprimer les fiches d'une moisson

A partir de l'onglet « Récoltes », sur la ligne correspondant à la moisson à supprimer, cliquer sur l'icône « poubelle »  pour supprimer les fiches correspondant à cette définition de moisson.

Cette action peut-être utile dans le cas où on ne veut plus des fiches de cette définition, ou dans le cas où l'on veut recommencer une moisson depuis le début. Dans ce dernier cas, il faut supprimer la récolte, puis relancer la moisson.

Tâches programmées

Cliquer sur l'onglet « Tâches programmées », la liste des moissons programmées en cours s'affiche, avec pour chacune les renseignements suivants :

- **JobName** : nom de la tâche, qui reprend l'identifiant de la moisson suivi d'un chiffre,
- **NextFire** : date prévue pour la prochaine exécution de la tâche,
- **PreviousFire** : dernière date d'exécution de la tâche,
- **TriggerName** : nom du déclencheur de la tâche.

Liste des programmations de moissons

JobName	NextFire	PreviousFire	TriggerName
CEL_Set_SH50	05/04/09 00:10		harvests-trigger.CEL_Set_SH50

Rapports

Cet onglet permet de consulter les rapports concernant les moissons qui ont été exécutées.

Rapports de moissons

Identifiant		Dernière moisson :									
CEL_Set_SH5		2009-03-10	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Date</th> <th>Documents ajoutés/mis à jour</th> <th>Documents supprimés</th> <th>Rapport</th> </tr> </thead> <tbody> <tr> <td>:</td> <td>:</td> <td>:</td> <td>:</td> </tr> </tbody> </table>	Date	Documents ajoutés/mis à jour	Documents supprimés	Rapport	:	:	:	:
Date	Documents ajoutés/mis à jour	Documents supprimés		Rapport							
:	:	:	:								
UNIT_Set-Informatique											

Pour chaque définition, on peut voir simplement le dernier rapport (icône œil de gauche), ou l'ensemble des rapports depuis la première moisson (icône œil de droite).

Lien vers le dernier rapport de moisson

Identifiant		Dernière moisson :	
CEL_Set_SH5		2009-03-10	
UNIT_Set-Informatique			

Lien vers l'ensemble des rapports de moisson

Chaque définition de moisson possède, pour chaque occurrence de moisson, une partie globale et une partie propre à chaque entrepôt défini pour cette moisson.

La partie globale contient les informations suivantes :

- la date d'exécution de la moisson
- un ensemble de compteurs pour les fiches ajoutés, mises à jour, ou supprimées, dont le détail est expliqué ci-après.

Ces compteurs sont donc eux-même subdivisés sur trois lignes :

- la date d'exécution de la moisson,
- le nombre de documents ajoutés ou mis à jour (total et par entrepôt),
- le nombre de documents supprimés (total et par entrepôt).

CEL_Set_SH5											
Date :	Documents ajoutés/mis à jour	Documents supprimés	Rapport :								
2009-03-10T12-19-23Z	11	0	harvesting CEL_Set_SH5 startedharvesting CEL_Set_SH5 finished : 11 documents added and 0 documents deleted in 3 s 797 ms <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <thead> <tr> <th style="width: 20%;">Nom de cet entrepôt</th> <th style="width: 15%;">Documents ajoutés/mis à jour</th> <th style="width: 15%;">Documents supprimés</th> <th style="width: 50%;">Durée</th> </tr> </thead> <tbody> <tr> <td>cel.archives-ouvertes.fr</td> <td>11</td> <td>0</td> <td>3 s 797 ms</td> </tr> </tbody> </table>	Nom de cet entrepôt	Documents ajoutés/mis à jour	Documents supprimés	Durée	cel.archives-ouvertes.fr	11	0	3 s 797 ms
Nom de cet entrepôt	Documents ajoutés/mis à jour	Documents supprimés	Durée								
cel.archives-ouvertes.fr	11	0	3 s 797 ms								

Tout ces détails qui peuvent paraître superflus, sont fournis pour aider à **diagnostiquer** les eventuels problèmes et afin de bien visualiser les choses.

Récupération

En cas de problème (index corrompu,...), il est possible de réindexer les fiches moissonnées, depuis l'onglet "Récupération", partie

Récupération.

Récupération

Ré-indexer toutes les fiches moissonnées

Attention : il faut bien s'assurer d'avoir supprimé l'index avant de lancer la restauration !

Tout ré-indexer

Ré-indexer toutes les fiches moissonnées depuis cette date (format AAAA-MM-JJ)

Tout ré-indexer depuis cette date

Attention : il faut bien s'assurer d'avoir restauré partiellement l'index avant de lancer la restauration !

Deux modes sont offerts.

- La ré-indexation de **toutes les fiches** : **supprimer l'index**, puis cliquer sur le bouton "Tout ré-indexer" ;
- La ré-indexation des **fiches moissonnées depuis une date donnée** (n'est possible que si des sauvegardes régulières de l'index ont été faites) : restaurer partiellement l'index (remplacer l'index par le backup approprié), puis cliquer sur le bouton "Tout ré-indexer depuis cette date".

Effectuer une veille sur les entrepôts à moissonner

Aujourd'hui, la vaste majorité des archives institutionnelles OAI concerne la production scientifique. En effet, le projet OAI (Open Archives Initiative : initiative des archives ouvertes) s'est initialement développé dans le champ de la recherche scientifique, afin de permettre l'échange de documents numérique au sein de la communauté des chercheurs.

C'est pourquoi, les répertoires qui existent en ligne proposent presque systématiquement des listes d'entrepôts relatifs à la production scientifique (thèses, documents scientifiques, publications,...).

Néanmoins, ces répertoires peuvent être un premier point de départ. On peut en citer quelques exemples, qui sont en anglais :

- le site officiel de l'OAI, Open Archives Initiative, propose un répertoire d'archives institutionnelles, sous la forme d'un tableau ; <http://www.openarchives.org/Register/BrowseSites> ; il propose aussi une liste d'agrégateurs de fiches de métadonnées, aussi sous la forme d'un tableau (ces agrégateurs fournissent en général des informations sur les entrepôts qu'ils moissonnent) ; <http://www.openarchives.org/service/listproviders.html> ;
- OAIster, agrégateur de fiches de métadonnées (15 millions à ce jour), propose un répertoire d'archives classées par ordre alphabétique ; <http://www.oaister.org/viewcolls.html> ;
- OpenDOAR : The Directory of Open Access Repositories, propose un répertoire d'archives en accès ouvert (open access) ; <http://www.openaccess.org> ;
- le site de Scientific Commons propose un répertoire d'archives classées par pays ; <http://en.scientificcommons.org/repository/overview>.

La veille sur les entrepôts de ressources pédagogiques se fera donc plutôt au cas par cas, en naviguant sur internet, à partir de mots-clés comme par exemple "entrepôt", "ressources pédagogiques", "OAI", "learning resource", "educational resource", "repository", etc. Aller voir du côté des communautés francophones (Belgique, Canada, Suisse,...) peut aussi être une autre source de pistes.

Tester un entrepôt avant de définir une moisson

Pourquoi tester un entrepôt et son contenu avant de définir une moisson ?

D'une part, pour vérifier que l'URL OAI mène bien vers un entrepôt OAI et que l'entrepôt est encore actif. D'autre part, pour analyser le contenu au regard du public et des besoins de son institution.

Il existe sur internet un outil qui permet l'analyse d'archives institutionnelles respectant le protocole OAI-PMH. Cet outil, "Open Archives Initiative - Repository Explorer", se trouve à l'adresse suivante : <http://re.cs.uct.ac.za/>. Il est hébergé et maintenu par l'Université du Cap en Afrique du Sud.

L'analyse se fait en utilisant l'une des six requêtes de l'OAI-PMH (pour plus d'informations se reporter à : <http://www.openarchives.org/OAI/openarchivesprotocol.html>). Certes sommaire, l'outil permet par exemple de connaître les formats de métadonnées gérés dans l'entrepôt, de savoir si l'entrepôt contient des ensembles ("sets" en anglais), d'accéder à la liste des fiches de métadonnées de l'entrepôt ou d'un ensemble (si il en existe).