
	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

# Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI


<b>Livrable du au titre du projet</b>	COCLICO
<b>Sous-projet</b>	WP2
<b>Tache</b>	2.6
<b>Livrable</b>	<b>L2.6.2</b>

<b>Rédacteur(s)</b>	<b>Vérificateur(s)</b>	<b>Approbateur(s)</b>
Luca Dini Hratchia Pélibossian Sebastien Dabet	O.Berger	

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0


<b>Documents applicables</b>
Annexe technique au projet COCLICO

<b>Documents de références (pour information)</b>
<a href="http://lucene.apache.org/solr/">http://lucene.apache.org/solr/</a>
<a href="http://lucene.apache.org/solr/tutorial.html">http://lucene.apache.org/solr/tutorial.html</a>
<a href="http://www.openarchives.org/">http://www.openarchives.org/</a>
<a href="http://www.openarchives.org/OAI/openarchivesprotocol.html">http://www.openarchives.org/OAI/openarchivesprotocol.html</a>
<a href="http://www.dublincore.org/">http://www.dublincore.org/</a>
<a href="https://forge.projet-coclico.org/projects/wp3/">https://forge.projet-coclico.org/projects/wp3/</a>
<a href="http://www.projet-coclico.org/index.php/DESCRIPTION_TECHNIQUE_DU_PROJET">http://www.projet-coclico.org/index.php/DESCRIPTION_TECHNIQUE_DU_PROJET</a>
<a href="https://forge.projet-coclico.org/projects/wp2/">https://forge.projet-coclico.org/projects/wp2/</a>

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0


## Gestion des versions

<b>N° de version</b>	<b>Date</b>	<b>Auteurs</b>	<b>Modification apportées</b>
<b>0.01</b>	20/05/10	Pélibossian	Première version
<b>1.0</b>	02/10/10	Pélibossian	Seconde version

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

## Sommaire

1	Objectif.....	5
2	Outils open source utilisés.....	6
3	Composant d'habilitation d'accès aux contenus des forges.....	7
4	Bibliographie .....	9

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

# 1 Objectif


Ce document présente un des résultat du sous-projet WP2 « *Interopérabilité et échange de données* ». Il fait un compte-rendu des activités de réalisation effectuées pour le livrable logiciel LI-2.6.2 du projet COCLICO « Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI ».

## Contexte

Ce document est le résultat de travaux effectués dans la tâche 6 « Fonctionnalités pour la fédération de forges » du sous-projet 2 « Interopérabilité et échange de données ».

Nous proposons un composant basé sur l'OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting) qui est un protocole d'échange et de transfert de données [OAI-PMH]. OAI-PMH est utilisé au niveau mondial par les acteurs du monde des archives électroniques ouvertes pour échanger des métadonnées.


L'architecture l'OAI-PMH centralise les métadonnées décrivant différentes ressources, en laissant ces ressources à leur emplacement initial. Par conséquent la publication d'information pour les utilisateurs sera effectué en local, nous pensons que c'est un avantage très appréciable. Il y a également dans le schéma proposé, une procédure de collecte d'informations.

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

## 2 Outils open source utilisés

L'outil réalisé, « CeliOAIHarvester », implémente un composant d'habilitation d'accès aux contenus des forges aux harvesters OAI. Il est réalisé en java et se base sur les outils open sources **SOLR et RVOT**. **Solr** [SOLR] est utilisé pour stocker les métadonnées de forges dans son index. Solr est un serveur de recherche pour entreprises, open source et basé sur la librairie Java de recherche Apache Lucene, avec des APIs XML/HTTP, le support du cache, de la réplication, et une interface d'administration Web.

L'outil **Rapid Visual OAI Tool** (RVOT) est utilisé pour construire graphiquement un entrepôt OAI-PMH à partir d'une collection de fichiers [RVOT]. Les dossiers de la collection d'origine peuvent être dans l'un des formats acceptables. Les formats pris en charge actuellement sont RFC1807, un sous-ensemble des formats Marc & COSATI, qui sont des formats de méta-données pour les bibliographies et la documentation technique. RVOT contribue à définir la correspondance visuelle à partir d'un format natif au format oai\_dc, et une fois cela fait l'outil peut répondre aux requêtes OAI-PMH. L'outil est autonome, il est livré avec un serveur web léger et gestionnaire de requête OAI-PMH et est écrit en Java.

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

### 3 Composant d'habilitation d'accès aux contenus des forges

OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) est un protocole d'échange et de transfert de données. Il vise à rendre partageable dans le monde entier des descriptions de documents (métadonnées) et à les rendre accessibles, sans les dupliquer ni modifier leur localisation d'origine. Il aide également à la synchronisation des différentes sources de données.

L'OAI-PMH définit trois acteurs différents dans une architecture de l'information basée sur les métadonnées :

- **Le fournisseur de données (ou fournisseur de contenus - *data provider*)**

Le fournisseur de données définit les données exposées, le(s) schéma(s) selon lesquels les données sont mises à disposition, et les rend accessibles sur un **entrepôt OAI** qu'il maintient à jour.

- **Le fournisseur de service (ou *service provider*)**

Le fournisseur de services lance un programme appelé **moissonneur** pour envoyer une requête à un fournisseur de données et en collecter les métadonnées. Le fournisseur de services traite les métadonnées qu'il a rassemblées et offre un service basé sur ces métadonnées. Par exemple une création d'un portail, un service de création et de mise à jour de catalogues et de bibliothèque des données, un service d'indexation et de recherche fédérée, un service d'accès et de visualisation des données.

- **L'agrégateur**

L'agrégateur est un troisième type d'acteur qui peut, dans certains cas, s'ajouter à cette configuration : il fournit des données intermédiaires. Il rassemble les métadonnées provenant de plusieurs fournisseurs de données et les rend accessibles dans un entrepôt OAI, éventuellement après les avoir retraitées. Les agrégateurs peuvent garantir des métadonnées de meilleure qualité, assurer un stockage commun, engager un travail de normalisation.

Voici un schéma correspondant à une telle architecture :

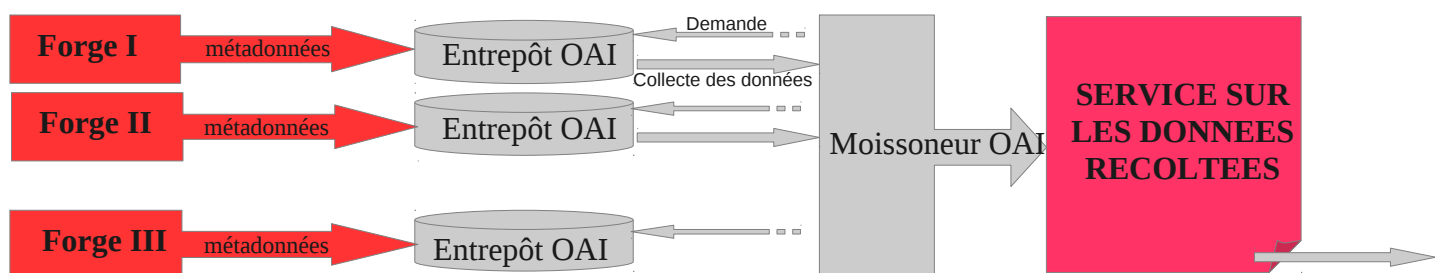



Schéma 1: Architecture OAI-PMH pour un portail de fédération de forges

Notre composant (en rouge sur le schéma) permet de fournir les données grâce aux index de **SOLR** de chaque Forge et à l'outil **RVOT** qui offre l'accès à ces données via serveur HTTP, dans le format compatible OAI PMH.

Le moissonnage des données est assuré grâce à **CeliOAIHarvester** qui récolte les données des serveurs HTTP mis en place par RVOT et les stocke sur serveur SOLR.

Les données sont alors consultable via un portail basé sur vuFind (cf. LI-2.6.3).

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0


### 3 Installation

#### Pré-requis

1. Windows 2000/NT/XP, Linux or Unix.
2. Java (JDK 1.6 and above) installed and CLASSPATH set appropriately.

#### Procédure d'installation

1. Télécharger "CeliOAIHarvester" qui se trouve dans le référentiel Bzr <http://forge.projet-coclico.org/scm/loggerhead/wp2/ce/i/files>.
2. Décompresser "CeliOAIHarvester".
3. Exécuter `lucene-index-to-RFC_1807.bat` pour créer les documents Lucene dans format RFC\_1807.
4. Télécharger SOLR from <http://lucene.apache.org/solr/>
5. Lancer Solr server pour OAI harvesting avec `schema.xml` qui se trouve dans `data/solr/schema.xml`.
6. Télécharger "Rapid Visual OAI Tool" Package qui se trouve à <http://rvot.sourceforge.net/>.
7. Décompresser "RapidVisualOAITool" .
8. Lancer "RapidVisualOAITool", qui transforme RFC\_1807 format dans format DC et lance repository http serveur à partir de "RapidVisualOAITool"
9. Modifier `conf/harvesting.xml`.
10. Lancer serveur Solr pour OAI harvesting.
11. Lancer `celi-oai-harvest.bat`. Il récolte les metadonnées dans SOLR.

	Titre du document : <b>Composant d'habilitation d'accès aux contenus des forges aux harvesters OAI</b>
	Référence : <b>L2.6.2</b>
	Version du 1.0

## 4 *Bibliographie*

[LI_2.6.1]	Livrable COCLICO : « Étude format d'échange compatible OAI-PMH»
[OAI-PMH]	OAI-PMH: <a href="http://www.openarchives.org/pmh/">http://www.openarchives.org/pmh/</a>
[RVOT]	RVOT: <a href="http://rvot.sourceforge.net/">http://rvot.sourceforge.net/</a>
[SOLR]	SOLR: <a href="http://lucene.apache.org/solr/">http://lucene.apache.org/solr/</a>