



ORI-OAI - Outil de Référencement et d'Indexation pour un réseau de portails OAI

Yohan Colmant, Raymond Bourges, François Jannin

► To cite this version:

Yohan Colmant, Raymond Bourges, François Jannin. ORI-OAI - Outil de Référencement et d'Indexation pour un réseau de portails OAI. JRES (Journées réseaux de l'enseignement et de la recherche) 2007, Renater, Nov 2007, Strasbourg, France. <hal-04802855v2>

HAL Id: hal-04802855

<https://hal.science/hal-04802855v2>

Submitted on 29 Nov 2024

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.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

ORI-OAI (Outil de Référencement et d'Indexation pour un réseau de portails OAI)

Raymond Bourges

CRI - Université de Rennes 1

raymond.bourges@univ-rennes1.fr

Yohan Colmant

Service Informatique - Université de Valenciennes

yohan.colmant@univ-valenciennes.fr

François Jannin

CRI - ENSEEIHT Toulouse

francois.jannin@enseeiht.fr

Résumé

Le processus de partage des ressources numériques s'inscrit dans le grand mouvement de l'accès ouvert aux ressources numériques et il s'approprie, naturellement, les mécanismes de mutualisation mis en place par ce mouvement mondial. C'est ainsi que nos portails doivent aujourd'hui être conformes au protocole d'échanges de données OAI-PMH pour pouvoir répondre aux exigences des nouvelles formes de mutualisation. Le Ministère de l'Education Nationale français (projet inter-UNT/SDTICE) soutient activement cette démarche initiée dans le consortium ESUP Portail et adoptée par les Universités Numériques Thématiques. Ce document présente le système ORI-OAI, libre, composé de procédures de travail et de composants logiciels, permettant de répondre au besoin des établissements pour le référencement, l'indexation et le partage de leurs ressources.

Mots clefs

Référencement, indexation, partage, documents numériques, OAI-PMH, fédération d'identités

1 Introduction

Le projet ORI-OAI [1] vise la mise en place d'un système ouvert, en open source, libre, permettant :

- de gérer toutes les ressources numériques produites par les établissements universitaires,
- de les partager avec d'autres établissements,
- de les valoriser par une indexation de qualité,
- de les rendre accessibles, à distance et selon les droits définis, dans des interfaces ergonomiques.

En effet, dans les établissements universitaires, l'explosion de la création numérique institutionnelle amène à se poser la question d'une gestion coordonnée et globale de cette production pour en assurer l'accès, sur la durée.

Plusieurs acteurs des établissements sont impliqués dans la production (enseignants, chercheurs, étudiants, le service

commun de documentation, le service TICE¹, ...) et plusieurs autres sont impliqués dans la valorisation et la gestion pérenne de ces contenus numériques (service commun de documentation, la direction des systèmes d'information, ...).

Le développement des UNT (Universités Numériques Thématiques) et des UNR (Universités Numériques en Région) pose la question du partage de ces ressources, dans des systèmes interopérables, accessibles depuis les ENT² des établissements et au sein des UNT et des UNR.

1.1 Historique

ORI-OAI découle de la réflexion de plusieurs initiatives sur la gestion des documents numériques :

- SYNAPSE de l'INSA de Lyon,
- inJAC du consortium ESUP-Portail [2],
- portail de ressources pédagogiques de l'UNT UNIT.

Le projet ORI-OAI dispose de financements publics issus, notamment, de la sous-direction des bibliothèques de la direction de l'enseignement supérieur et de la sous-direction des TICE de la direction technique du Ministère de l'Education nationale français.

2 Contour fonctionnel

2.1 Les acteurs

Le but du système ORI-OAI est de permettre aux utilisateurs de rechercher des ressources numériques en provenance d'un établissement, ou d'un ensemble d'établissements. Il permet ensuite l'accès à ces ressources de façon libre ou contrôlée.

Au-delà des acteurs qui accèdent en consultation aux ressources, le système est prévu pour permettre à un grand nombre de personnes de publier des ressources. Il peut

¹ Technologies de l'Information et de la Communication pour l'Education

² Espace Numérique de Travail : Portail WEB permettant l'accès à des services en relation avec le système d'information de l'établissement

s'agir d'enseignants, de chercheurs, de personnels administratifs ou même potentiellement d'étudiants.

De plus, la saisie des informations d'indexation plus complexes peut être laissée à un service spécialisé comme, par exemple, les services communs de documentation.

D'une façon générale le système offre une grande flexibilité de configuration afin de laisser le choix à l'établissement de son organisation quant aux processus de publication, de référencement et d'indexation des ressources.

2.2 Les ressources numériques

Le système ORI-OAI a été, depuis le début de sa conception, pensé pour travailler sur différents types de ressources numériques. Chaque ressource pourra, suivant son type, être référencée en utilisant un format de métadonnées adapté.

Les ressources peuvent, par exemple, être des ressources pédagogiques (polycopiés, énoncés et corrigés d'exercices, etc.), des travaux d'étudiants (rapports de stage ou de projets, synthèses bibliographiques, etc.), des travaux de recherche (publications, rapports techniques, mémoires de master, mémoires de thèses, etc.) ou des ressources documentaires éditoriales acquises par l'établissement (périodiques électroniques, livres électroniques, etc.).

2.3 Les fonctionnalités

Le système ORI-OAI est développé afin de répondre à de nombreux objectifs parmi lesquels on peut citer les éléments clés suivants :

Référentiel unique, pour les ressources numériques de l'établissement. Le référentiel ne se substitue pas aux diverses plateformes qui peuvent utiliser et publier ces mêmes ressources (plateformes pédagogiques, sites web des laboratoires ...), mais il gère la forme canonique de la ressource et permet éventuellement un contrôle d'accès au sein d'une fédération d'identités,

Système de recherche avancé, multicritères,

Accès thématiques aux ressources, selon des classifications simplifiées qui exploitent la classification Dewey [3],

Système de gestion et de publication des ressources numériques par une publication web avec gestion des droits d'accès ; description des ressources selon les normes LOM [4], TEF [5], Dublin Core [6], etc. en relation avec les autres systèmes documentaires pour le partage des tables d'autorité (SUDOC [7], STAR [8] par exemple) ; Indexation selon les classifications en usage dans les bibliothèques universitaires (Dewey par exemple) et exploitées par les classifications spécifiques des UNT ; archivage des ressources numériques,

Système de production impliquant les acteurs concernés, dans des procédures élaborées, matérialisées par des *workflows* ; gestion des versions et accès aux versions natives des documents pour leurs créateurs,

Système de partage, fondé sur l'échange de métadonnées selon le protocole OAI-PMH [9], permettant de fonctionner au sein d'une communauté constituée (UNT par exemple) en réseau de portails,

Système open-source, libre documenté et pouvant être installé simplement.

3 Concepts

3.1 Le concept de métadonnées

Les métadonnées représentent l'ensemble des informations externes associées à une ressource. Cet ensemble permet de définir pour chaque ressource un contexte d'utilisation, et un jeu de relations entre les ressources. L'exemple simple est celui du nom d'un auteur, qui permet d'appréhender l'ensemble des ressources qu'il a produites, ou encore celui d'un code issu d'une taxonomie, qui offre une vue sémantiquement structurée d'un ensemble de ressources.

Le concept de métadonnée définit, à travers des formats concrets forgés au fil des pratiques (Dublin Core, LOM, CDM [10], TEF [11], MARC [12], ETDMS [13], etc.), un regroupement déterminé de données qui peuvent être semi-structurées (grâce au langage de balise XML). Ce regroupement agrège et anticipe les besoins d'une communauté autour de l'ensemble de ses activités, et assure une sémantique forte grâce à l'utilisation de vocabulaires fermés, standardisés, d'où il tire une très grande efficacité grâce à la qualité de l'information ainsi produite. Ces vocabulaires fermés sont définis à l'intérieur des schémas XML qui valident les métadonnées pour un format donné.

Maintenant, qu'entend-on par entrepôt de données ? C'est un système capable d'associer un ou plusieurs documents à un jeu de métadonnées, afin de constituer un référentiel pour un ensemble ouvert d'utilisations, dans la limite définie par le format utilisé.

Ainsi, le format Dublin Core non qualifié définit un jeu de 15 métadonnées à vocation très généraliste, facultatives et répétables, qui peut suffire à représenter des relations simples mais non ambiguës dans un ensemble de ressources : un master peut communément être décrit par son auteur, la date de sa création et publication, un code le situant dans l'ensemble des domaines de l'école où il est produit, et quelques autres métadonnées.

Plus complexe, le format LOM est le fruit de la conceptualisation de toutes les informations pertinentes pour décrire un objet pédagogique ; sa structure riche, détaille chacun des aspects par lesquels cet objet peut être appréhendé dans l'ensemble des activités dont il relève : aspect pédagogique pour la gestion scolaire et les conditions pédagogiques d'utilisation (niveau, type de destinataire, durée) qui en détermine l'auditoire, aspect technique pour ses conditions matérielles d'utilisation, aspect juridique pour les conditions légales, aspect du cycle de vie pour ses évolutions, ses versions... Ce format va jusqu'à définir un aspect de méta-métadonnées, pour décrire des informations sur les métadonnées elles-mêmes !

Pour résumer, les métadonnées forment donc la base d'un référencement de qualité, en décrivant de façon adaptée chaque ressource, lui donnant ainsi sa place dans un réseau de relations au sein d'une activité, ce qui permet aux acteurs de cette activité, producteurs comme consommateurs, de retrouver et de manipuler chaque ressource le plus aisément possible.

3.2 Le protocole OAI-PMH

Le protocole OAI-PMH [9] a été créé par l'organisation Open Archive Initiative [14], dédiée aux archives ouvertes et à la mise en œuvre pratique du partage de ressources. Chaque ressource est à cet effet représentée par un **enregistrement** OAI, composé d'un entête OAI et d'un jeu de métadonnées comme précédemment décrit. L'entête contient des informations utilisées par le protocole OAI pour les critères de sélection tels que date de modification, ensembles auxquels l'enregistrement appartient, ou encore son statut supprimé. Chaque format de métadonnées doit être décrit par un schéma XSD dont l'espace de nom et l'adresse internet du fichier sont définis dans les réponses OAI, ce qui rend éligible comme format OAI n'importe quel format décrit par un tel schéma.

A l'aide d'une syntaxe composée de six verbes, simples mais efficaces, OAI-PMH permet de disséminer ces enregistrements par un mécanisme faisant intervenir un fournisseur de service, le **moissonneur**, et un ou plusieurs **entrepôts**, chacun associé à un fournisseur de données.

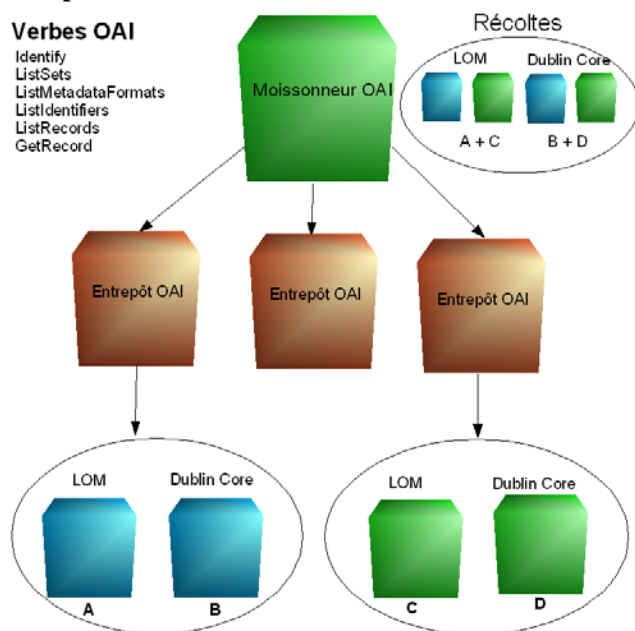


Figure 1 – Schéma OAI

Comme le montre la Figure 1, le mécanisme général fait donc intervenir un moissonneur OAI qui récolte de façon sélective des fiches stockées dans plusieurs entrepôts OAI, selon différents **critères de sélection**. Les critères de sélection d'une moisson doivent définir au minimum le **type de format** de métadonnées désiré (comme par exemple ici LOM et Dublin Core), et optionnellement un **ensemble** qui cible plus précisément un domaine ou toute forme de regroupement pertinent pour une utilisation particulière, ainsi que des **dates** de début et de fin. La notion d'ensemble (set OAI) relève d'une fonctionnalité assumée de façon interne par l'entrepôt, pour regrouper ensemble des enregistrements dont le format de métadonnées commun possède des éléments permettant une classification.

Les critères de date permettent d'opérer des moissons incrémentales, c'est-à-dire de ne pas récupérer à chaque

fois tous les enregistrements, mais seulement ceux qui ont été ajoutés, modifiés ou supprimés depuis la dernière moisson, ceci afin que le moissonneur puisse traiter correctement le cycle de vie des enregistrements récoltés.

Le premier verbe est **Identify**, qui fournit une fiche d'identité de l'entrepôt, pouvant comporter une description des contenus disponibles ou des liens sur d'autres entrepôts « amis ». Deux verbes, **ListSets** et **ListMetadataFormats**, fournissent respectivement une liste des ensembles disponibles et une liste des formats supportés. L'accès aux fiches se fait par les verbes **ListRecords** et **ListIdentifiers**, le deuxième ne renvoyant que l'entête muni de l'identifiant de l'enregistrement. Grâce à cet identifiant, on peut récupérer une seule fiche individuellement avec le sixième verbe **GetRecord**.

Les verbes OAI qui retournent des listes utilisent un jeton de continuation (*resumption token*) pour découper les listes trop longue en plusieurs morceaux, plus facilement manipulables dans les flux sur le réseau.

3.3 Fédération d'identités avec Shibboleth

L'objectif d'une fédération d'identités est de faciliter le partage de ressources numériques en ligne entre établissements en interconnectant leurs services d'authentification. Il devient possible d'ouvrir l'accès à une ressource (pédagogique, scientifique, éditoriale, application métier, etc.) à une population identifiée, sans devoir gérer localement l'enregistrement des utilisateurs.

La fédération d'identités concrétise, pour un groupement d'établissements, l'interconnexion de leurs services d'authentification et l'utilisation d'un ensemble commun d'attributs utilisateurs. Un établissement qui gère un ensemble d'utilisateurs est appelé fournisseur d'identités. Un fournisseur de services est une entité - établissement, administration, société privée - qui propose une ressource numérique en ligne au sein de la fédération. Techniquement, les relations de confiance entre les membres d'une fédération reposent sur des certificats électroniques et une définition partagée des différents fournisseurs d'identités et de services. En outre, la confiance s'établit entre les participants de la fédération au travers d'une convention.

Un même établissement peut participer à plusieurs fédérations et gérer des partenariats de manière bilatérale. Il peut également jouer à la fois le rôle de fournisseur d'identités et de fournisseur de services.

En France, pour l'enseignement supérieur, les règles d'usage et l'assistance technique sur la fédération d'identités sont fournies par le CRU [15]. L'implémentation technique choisie par le CRU comme par le projet ORI-OAI est celle développée dans le contexte universitaire de Internet2, à savoir : Shibboleth [16].

Le schéma suivant décrit de façon très simplifiée le mécanisme de fédération d'identités :

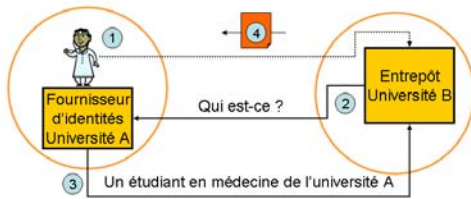


Figure 2 – Le mécanisme Shibboleth

1. L'utilisateur tente d'accéder au document dans un autre établissement
2. L'établissement B demande des informations sur l'émetteur de la requête au fournisseur d'identités de l'établissement d'appartenance de l'utilisateur
3. L'établissement A fournit des attributs sur l'utilisateur courant au fournisseur de services de l'établissement B
4. Le fournisseur de services de l'établissement B vérifie ces attributs afin de renvoyer le document ou pas.

3.4 Exemples de contextes de mise en application

Un moissonneur permet de rassembler, en un lieu unique, un ensemble de métadonnées provenant de différents entrepôts. Il est alors possible d'indexer et de proposer une interface de recherche sur ces métadonnées afin de localiser et donner accès à la ressource afférente.

Il est possible d'imaginer de nombreuses mises en application. Nous allons en décrire deux à titre d'exemple :

- Un internaute consulte le site web d'un consortium (exemple : regroupement d'universités sur une thématique ou sur une région) à la recherche de ressources *a priori* libres.
- Un utilisateur d'un établissement, dans le cadre des services proposés par son ENT, recherche des ressources dans son propre établissement et les établissements partenaires. Dans ce cas, les ressources peuvent être soumises à un contrôle d'accès.

3.4.1 Site Web d'un consortium

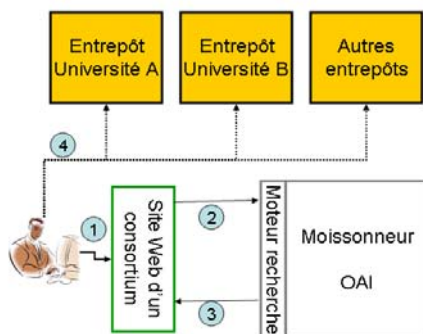


Figure 3 – Le contexte d'un site Web

1. L'utilisateur accède au site web du consortium,

2. Il formule une recherche sur le moteur de recherche du consortium,
3. Il obtient des références vers des ressources,
4. Il accède à ces ressources dans les différents entrepôts présents dans les établissements membres du consortium.

3.4.2 Accès depuis un ENT d'établissement

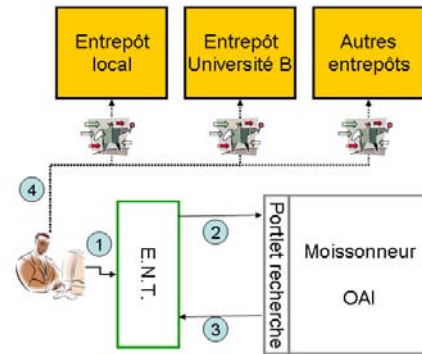


Figure 4 – Le contexte d'un ENT

1. L'utilisateur s'identifie pour accéder à son portail de services,
2. Il formule une recherche sur le portlet (JSR 168) [17] de recherche ORI-OAI installé dans son établissement,
3. Il obtient des références vers des ressources,
4. Il accède à ces ressources dans l'entrepôt de son établissement ou des établissements partenaires. Si la ressource accédée nécessite un contrôle d'accès les mécanismes de la fédération d'identités sont utilisés de façon transparente pour l'utilisateur (il n'a pas besoin de s'identifier à nouveau).

4 Mise en œuvre

4.1 L'architecture globale du projet

L'architecture du projet ORI-OAI se présente sous la forme de sept composants développés indépendamment. Chacun de ces composants a un rôle bien défini dans le système et communique avec les autres au travers de services exposés en frontal de chaque composant. La technologie choisie pour rendre possible le dialogue entre tous les modules est les Web services [18].

Ce choix permet une grande souplesse dans d'éventuelles déclinaisons d'architectures et de langages de programmation. En effet, cette architecture technique a été notamment pensée pour permettre à des logiciels extérieurs de dialoguer avec différents éléments du système. Nous pouvons par exemple prendre le cas d'une bibliothèque qui souhaiterait faire des recherches de documents depuis son portail documentaire sans passer par l'interface de recherche ORI-OAI-search. Il est donc possible qu'il y ait une connexion directe entre leur logiciel et l'index de recherche ORI-OAI-indexing.

Notons que ORI-OAI est développé en Open Source sous licence GPL (General Public License) [19].

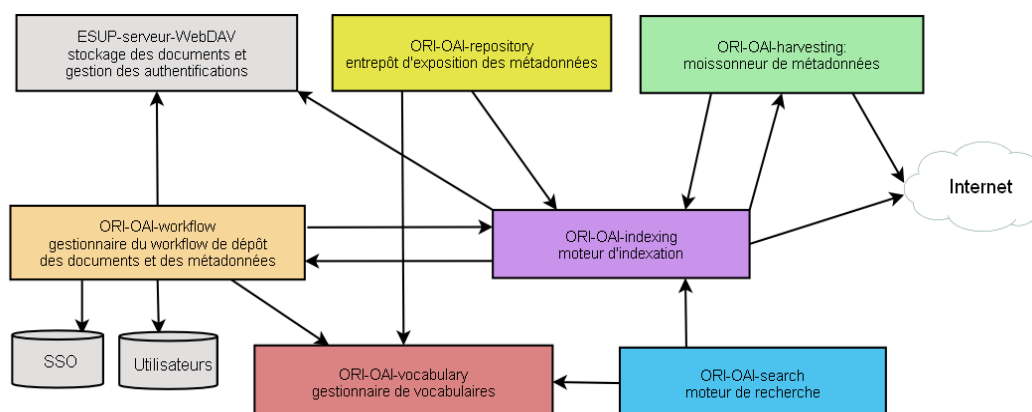


Figure 5 – L'architecture globale

4.2 Les différents modules

Nous avons vu que tous les modules sont développés indépendamment les uns des autres, cependant, ils reposent sur certaines technologies de base commune. En effet, le langage de développement Java [20] ainsi que l'environnement de développement Spring [21] représentent les technologies de base de chaque composant. De plus, toutes les communications par Web services sont implémentées via XFire [22].

Voyons maintenant les fonctionnalités de chaque module ainsi que les technologies qu'ils emploient. Soulignons que toutes ces technologies sont des projets Open Source.

ORI-OAI-workflow

Ce composant est utilisé pour le dépôt des documents dans le système, l'affectation des droits d'accès aux ressources, ainsi que la saisie des métadonnées. Les technologies utilisées par ce moteur de dépôt sont diverses et permettent un paramétrage très fin, complet et puissant en fonction des besoins de l'établissement ou du consortium qui le met en place.

Le moteur de *workflow* OsWorkflow [23] permet une configuration avancée de toutes les actions effectuées à chacune des étapes du dépôt. En effet, par configuration, il peut répondre aux besoins des établissements exigeant du déposant la saisie de tous les champs de métadonnées, mais aussi des établissements plus structurés où la saisie des métadonnées s'effectue en différentes étapes, par différents intervenants. Chaque établissement peut ainsi configurer le module suivant ses besoins.

Les formulaires de saisie des métadonnées sont eux aussi configurables. La technologie XForms [24] associée au moteur de transformation Orbeon [25] offre des formulaires de saisie dynamiques en fonction de fichiers de configuration écrits en XML. Le support des différents formats de métadonnées repose sur la possibilité d'enrichir le module ORI-OAI-workflow de nouvelles configurations XForms.

Les autres technologies utilisées dans ce module sont Acegi Security [26] pour gérer toutes les sécurités d'accès à l'application, eXist [27] comme base de données XML Open Source pour le stockage des fiches de métadonnées,

Hibernate [28] pour rendre les accès à la base de données relationnelle transparents et enfin JSF [29] pour la gestion des interfaces graphiques.

ORI-OAI-indexing

Une fois le dépôt de ressources et la saisie de métadonnées validés, ceux-ci sont indexés par le module ORI-OAI-indexing. Ce module a pour rôle l'indexation des fiches de métadonnées ainsi que des documents associés.

Pour cela, il utilise le moteur d'indexation Lucene [30]. Celui-ci permet l'indexation de différentes sources offrant une recherche puissante et rapide en se reposant sur différents analyseurs. L'analyseur de la langue française permettra notamment la gestion des verbes conjugués, des pluriels ou encore des accents et caractères spéciaux. Un système de pondération permet aussi de rendre une métadonnée plus pertinente qu'une autre. Par exemple, on préférera retrouver en premier les documents dont l'élément recherché se trouve dans le titre plutôt que dans la description.

Lius [31] est un environnement d'indexation basé sur le projet Lucene. Il permet une indexation de différents formats de fichiers comme XML, PDF, OpenOffice, ZIP, MP3, etc. Il est utilisé dans notre projet pour offrir une configuration avancée des champs à indexer dans les différents formats de fiches de métadonnées XML et pour indexer les documents associés en plein texte.

En plus de l'aspect indexation, ORI-OAI-indexing offre un service de recherche de documents via Web service en se reposant sur la syntaxe des requêtes Lucene. Il est utilisé par différents composants dans le système.

ORI-OAI-repository

Nous avons vu dans un chapitre précédent le protocole OAI-PMH. Le module ORI-OAI-repository se charge, via ce protocole, de l'exposition des fiches de métadonnées saisies dans le module ORI-OAI-workflow. Utilisant le logiciel OAICat [32], il expose les fiches dans le but d'être moissonnées par tout moissonneur OAI.

Ce module gère également le concept de « sets OAI-PMH ». Cet aspect du protocole OAI permet d'exposer les fiches de métadonnées sous forme d'ensembles distincts. Ces ensembles sont souvent liés à une thématique particulière. Nous pouvons par exemple identifier

l'ensemble de toutes les fiches pédagogiques au format LOM associées aux notions de mathématiques. Pour identifier les fiches correspondant aux différents ensembles, le module ORI-OAI-repository construit des requêtes suivant les critères associés aux différents sets et les envoie au module ORI-OAI-indexing.

ORI-OAI-harvesting

Ce composant du système correspond au moissonneur OAI-PMH. Utilisant le logiciel OAIHarvester2 [33], il permet le moissonnage de fiches de métadonnées sur tout entrepôt OAI. Les fiches moissonnées sont alors stockées localement dans une base de données XML eXist.

Tout comme le moteur de *workflow*, ce module fournit toutes les fiches de métadonnées moissonnées au moteur d'indexation dans le but d'être indexées. L'index de recherche Lucene est alors composé des documents locaux, mais aussi des fiches de métadonnées moissonnées par ce module.

Via une interface graphique conviviale, l'administrateur du système peut programmer les différentes moissons qui seront lancées par le gestionnaire de tâches Quartz [34].

ORI-OAI-search

Ce module offre une interface graphique pour la recherche de documents dans le système. Dialoguant avec le module ORI-OAI-indexing, il génère des requêtes dans la syntaxe Lucene et affiche les documents retrouvés.

Ce composant est entièrement configurable en ce qui concerne les formats de documents que l'on souhaite rechercher, et les types de recherche que l'on veut proposer à l'utilisateur. Il existe trois types de recherche :

- Par date : on propose à l'utilisateur de rechercher des documents suivant leur date de création, modification, etc. Ce type est utilisé pour afficher les nouveautés.
- Simple ou Avancée : il est possible de configurer différents formulaires de recherche avec des critères plus ou moins avancés. On pourra proposer par exemple un formulaire de recherche simple composé d'un seul champ permettant une recherche sur le document complet et les métadonnées associées, ou encore un formulaire de recherche avancée où l'on pourra rechercher indépendamment sur chaque métadonnée d'un format donné.
- Thématique : dans ce type de recherche, on ne demande aucune saisie à l'utilisateur. Elle est mise en place pour faire des recherches suivant des classifications de documents, des auteurs, des mots-clefs, etc. Elle peut être sollicitée par les utilisateurs souhaitant découvrir les documents référencés dans le système n'ayant aucun critère de recherche particulier. Par ce type de recherche, on guide également l'utilisateur dans ses recherches en ne proposant que les valeurs réellement indexées comme, par exemple, dans le cas de la recherche par mots-clefs.

Notons également que le module peut être décliné en deux versions : *servlet* pour une installation standard sur un

serveur Web et *portlet* pour une intégration dans un Environnement Numérique de Travail.

ORI-OAI-vocabulary

Le composant ORI-OAI-vocabulary est celui qui gère les vocabulaires utilisés dans différents modules. On entend par vocabulaire un ensemble fermé de valeurs disponibles pour un critère donné.

Les vocabulaires peuvent être statiques et configurés via des fichiers XML comme, par exemple, des classifications de documents ou des valeurs strictes de champs de métadonnées LOM. Ils peuvent aussi être dynamiques comme dans le cas de toutes les valeurs disponibles dans l'index d'une métadonnée précise. Par exemple, on pourra constituer dynamiquement, en interrogeant le module ORI-OAI-indexing, la liste des mots-clefs ou des auteurs qui ont déjà été saisis dans les documents pédagogiques.

Ces vocabulaires sont utilisés par :

- Le module ORI-OAI-workflow pour proposer des listes de valeurs lors de la saisie des métadonnées
- Le moteur de recherche ORI-OAI-search pour les recherches thématiques ou les valeurs disponibles pour certains champs de la recherche avancée
- L'entrepôt ORI-OAI-repository pour générer dynamiquement des sets OAI en fonction par exemple d'une thématique donnée.

ESUP-serveur-WebDAV

Initialement développé par le consortium ESUP Portail, ce serveur permet le stockage des documents en ligne en utilisant le protocole WebDAV [35]. Le socle de ce composant est le serveur Jakarta/Slide [36]. Via différentes méthodes du protocole WebDAV, il est possible de gérer des documents distants (dépôt, téléchargement, suppression, etc.). De plus, le protocole ACP (Access Control Protocol) offre une gestion dynamique des droits (lecture, écriture, etc.) sur les documents.

Les développements qui ont été faits autour de ce serveur sont :

- Le support d'authentifications via le SSO (*Single Sign-On*) CAS [37]
- La gestion de fédération d'identités par Shibboleth
- L'externalisation de la gestion des groupes d'utilisateurs
- Le support des quotas.

5 Conclusion

Actuellement, le projet ORI-OAI offre dans sa version 1 les fonctionnalités d'un outil de référencement pour des ressources liées à la recherche (comme les maîtrises, master, publications scientifiques), ainsi qu'à toutes les ressources pédagogiques, grâce aux jeux de métadonnées Dublin Core et le standard LOM. Avec une implémentation complète du protocole OAI PMH, ainsi que l'utilisation des Web Services, il pose les bases d'une large interopérabilité avec d'autres systèmes. Son système de gestion de *workflow* permet de configurer de façon très souple les rôles des différents intervenants lors de la production et la validation des métadonnées, dont la qualité

est encore renforcée par l'utilisation de vocabulaires standardisés et partagés par l'ensemble du système.

La composition modulaire d'ORI-OAI forme ainsi un noyau fonctionnel qui s'est voulu dès le départ très ouvert sur de nombreuses possibilités d'utilisations, rendant envisageable l'interopérabilité avec des systèmes nationaux comme HAL [38], STAR, etc., mais aussi avec des applications largement utilisées par la communauté de l'enseignement supérieur, telle que la plateforme pédagogique Moodle [39].

L'ambition que le Ministère de l'Education nationale et de la Recherche français et les différentes UNT et UNR impliquées ont placée dans ce projet est donc en voie d'être poursuivie à travers ces différents axes, afin d'aboutir dans un futur proche à une généralisation de son emploi dans une majorité d'établissements de l'enseignement supérieur, en France et pourquoi pas au-delà, dans des partenariats internationaux réunissant des universités de tout pays.

Bibliographie

- [1] <http://www.ori-oai.org/>
- [2] <http://www.esup-portail.org>
- [3] <http://www.oclc.org/dewey>
- [4] <http://ltsc.ieee.org/xsd/lomv1.0/lom.xsd>
- [5] <http://www.abes.fr/abes/documents/tef/index.html>
- [6] <http://dublincore.org>
- [7] <http://www.sudoc.abes.fr>
- [8] <http://www.abes.fr/abes/page,428,star.html>
- [9] <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [10] <http://cdm.utdanning.no/cdm>
- [11] <http://www.abes.fr/abes/documents/tef/index.html>
- [12] <http://www.loc.gov/marc>
- [13] <http://www.ndltd.org/standards/metadata/etd-ms-v1.01.html>
- [14] <http://www.openarchives.org>
- [15] <http://federation.cru.fr>
- [16] <http://shibboleth.internet2.edu/>
- [17] <http://jcp.org/en/jsr/detail?id=168>
- [18] <http://www.w3.org/2002/ws/>
- [19] <http://www.gnu.org/licenses/gpl.html>
- [20] <http://www.java.com/>
- [21] <http://www.springframework.org/>
- [22] <http://xfire.codehaus.org/>
- [23] <http://www.opensymphony.com/osworkflow>
- [24] <http://www.w3.org/MarkUp/Forms/>
- [25] <http://www.orbeon.com/>
- [26] <http://www.acegisecurity.org/>
- [27] <http://exist.sourceforge.net/>
- [28] <http://www.hibernate.org/>
- [29] <http://java.sun.com/javaee/jaserverfaces/>
- [30] <http://lucene.apache.org/java/docs/index.html>
- [31] <http://www.bibl.ulaval.ca/lius/>
- [32] <http://www.oclc.org/research/software/oai/cat.htm>
- [33] <http://www.oclc.org/research/software/oai/harvester2.htm>
- [34] <http://www.opensymphony.com/quartz/>
- [35] <http://www.webdav.org/>
- [36] <http://jakarta.apache.org/slide/>
- [37] <http://www.ja-sig.org/products/cas/>
- [38] <http://hal.archives-ouvertes.fr>
- [39] <http://moodle.org>

