

Un exemple de fédération de contenus : le projet ORI-OAI

Alain Kavenoky-UNIT et le groupe de développement ORI-OAI

alain@kavenoky.fr

www.unit.eu

www.ori-oai.org

ORI-OAI, Outil de Référencement et d'Indexation conforme au protocole OAI-PMH, est un système ouvert, en logiciel libre, permettant :

- de gérer et de publier tous les documents numériques produits 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.

Le développement ORI-OAI s'appuie sur les travaux de plusieurs groupes, consortiums et universités numériques.

Le protocole OAI (Open Archive Initiative) tire ses origines de l'apparition des bibliothèques de documents digitaux. La première initiative a été le serveur xxxx, devenu arXiv, créé par la communauté des physiciens des hautes énergies en 1991, puis étendu aux mathématiques et à l'informatique.

Des bibliothèques équivalentes sont ensuite apparues dans le domaine de l'économie, de l'archivage des thèses, chaque bibliothèque, conçue indépendamment des autres était dotée d'une interface informatique différente. L'utilisateur, intéressé par leur interrogation devait apprendre l'usage de chaque interface.

En juillet 1999, Paul Ginsparg, le créateur de arXiv, et quelques collègues ont appelé à la réunion d'une conférence pour définir un format standard de dépôt pour les documents archivés par leur auteur.

Il est rapidement apparu que l'interrogation, pour chaque requête, de nombreuses bibliothèques digitales était pénalisante, le temps de réponse étant imposé par la plus lente. L'idée de « moissonner » régulièrement un ensemble de bibliothèques permettait de proposer une solution efficace. Deux types de services ont été définis, les fournisseurs de contenus et les fournisseurs de services. Les premiers, en grand nombre et dispersés, archivent les contenus, les seconds, en petit nombre, agrègent les informations sur les meta-données des premiers et répondent aux requêtes des utilisateurs.

Le protocole OAI-PMH (Protocol for Meta-data Harvesting) a été conçu en s'appuyant sur le protocole de transport http et des échanges d'information en format XML. Le dialogue entre le moissonneur et les entrepôts est basé sur six « verbes » il permet au moissonneur :

- de vérifier si le service désigné est un répertoire OAI (**Identify**) ;
- de demander la liste des formats de méta données que l'entrepôt est capable de fournir (**ListMetadataformats**) ;
- d'obtenir, le cas échéant, la liste des ensembles (par exemple thématiques) formés par le fournisseur de données (**ListSets**) ;
- d'obtenir la liste des identifiants et des dates de dernière modification des items disponibles dans la base (**ListIdentifiers**) ;

- de collecter l'information d'un item spécifique (**GetRecord**) ;
- de récupérer en bloc l'ensemble des items d'un répertoire (**ListRecords**).

Lorsque le consortium UNIT s'est posé la question du développement d'une nouvelle version de son portail, il s'est très vite rendu compte qu'un portail unique, regroupant les contenus pédagogiques produits par les cinquante établissements membres poserait rapidement des problèmes de performances informatiques. Il a choisi de s'appuyer sur le protocole OAI-PMH. Les partenaires sont chargés d'exposer les méta données résultats de l'indexation des contenus disponibles et UNIT assure la moisson de tous les serveurs.

Cette question s'est posée, de la même manière, pour les autres UNT. UVED a rejoint très vite les réflexions d'UNIT et s'est impliqué dès le démarrage du projet. AUNEGE et UNISCIEL se sont ensuite joints en participant aux groupes de réflexion sur les portails communicants OAI. L'UMVF s'y intéresse et a lancé en groupe de travail dédié.

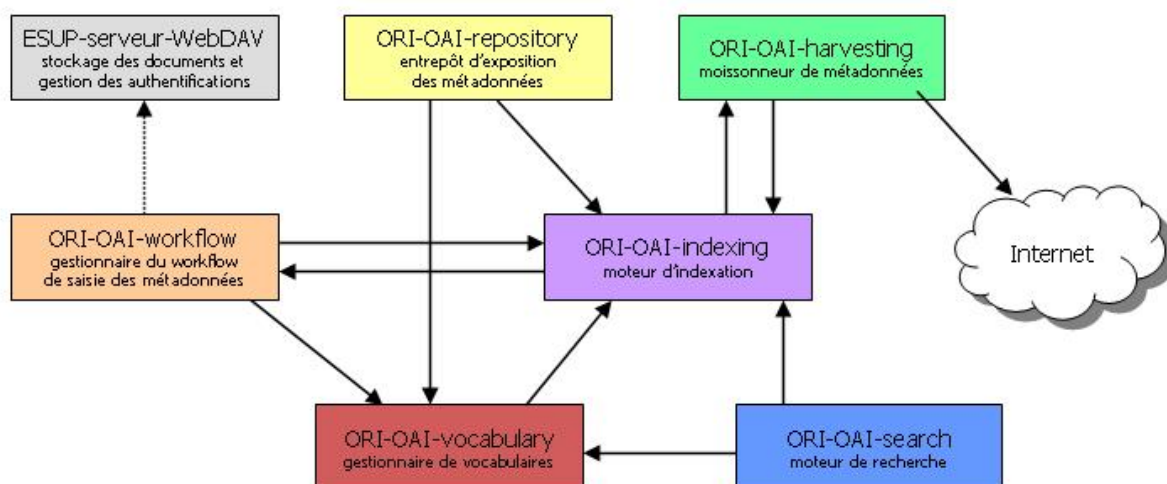
La pluridisciplinarité, étant à l'honneur parmi ses partenaires, UNIT a bien perçu l'intérêt d'accéder à certains cours d'économie d'AUNEGE qui utiliserait volontiers les contenus mathématiques d'UNISCIEL. La moisson, dans le futur doit concerner toutes les UNT.

Ce raisonnement a conduit la SDTICE à classer le projet OAI comme un projet transversal et à en assurer un financement direct.

L'INSA de Lyon, partenaire UNIT, qui travaille depuis plusieurs années au projet SYNAPSE de création d'une archive institutionnelle, rejoint le projet qui devient ORI-OAI en ajoutant le besoin d'un Outil de Référencement et d'Indexation pour tous types de ressources numériques pédagogiques, scientifiques ou administratives.

Les composants du projet

ORI-OAI est un projet transverse, inter-UNT, développé en collaboration par l'INPT, l'INSA de Lyon, les universités de Rennes 1 et de Valenciennes. Le logiciel produit est un logiciel libre, disponible pour tous, soumis à la licence GPL, basé sur une capitalisation des travaux menés par ESUP. Il est constitué de 7 briques logicielles indépendantes et interconnectables.



Le moteur de Workflow : ORI-OAI-workflow

Le moteur de Workflow gère le flux des données en local selon les procédures de travail définies par l'établissement. Il assure la coordination entre les différents acteurs : les auteurs, les documentalistes, etc. en permettant la répartition de la saisie des métadonnées d'un document. Différents formats de métadonnées peuvent être proposés, selon les documents gérés.

L'éditeur de métadonnées a été développé avec l'aide de la société ORBEON. Les fiches créées sont stockées sur une base de données.

L'entrepôt OAI : ORI-OAI-repository

Il expose un ensemble de fiches de métadonnées à la moisson au travers du protocole OAI-PMH. Il répond aux requêtes émises par tout moissonneur OAI. L'entrepôt peut afficher les différents formats de métadonnées présents dans l'application (formats de métadonnées moissonnées ou produites).

Le Moissonneur : ORI-OAI-harvesting

Le rôle de ce module est de moissonner les fiches d'un ensemble paramétrable d'entrepôts OAI via le protocole OAI-PMH. De même que dans le module workflow, ce module stocke les fiches moissonnées sur une base XML (eXist).

L'Indexeur : ORI-OAI-indexing

Il indexe les fiches moissonnées et produites, en utilisant le logiciel libre Lucene/Lius (Lucene Index Update and Search).

Le Moteur de Recherche : ORI-OAI-search

Le moteur de recherche, interfacé avec l'utilisateur, interroge les index des fiches pour indiquer à l'utilisateur les documents locaux et distants répondant aux critères demandés.

Le module Vocabulaire : ORI-OAI-vocabulary

Ce composant sert à répertorier et à gérer les différents vocabulaires et classifications de l'application. Il peut être utilisé pour la saisie de mot-clefs contrôlés, de noms d'institutions, de classifications disciplinaires, etc. Certains vocabulaires peuvent être gérés de manière globale au niveau d'une UNT, de plusieurs UNT, nationalement, etc. Ils doivent alors être cohérents pour toutes les entités concernées, tandis que d'autres vocabulaires peuvent être définis localement.

Le serveur WebDAV : ESUP-serveur-WebDAV

La gestion des contenus numériques (documents) est opérée par ce serveur. Il permet l'authentification SSO (single sign-on) et la fédération d'identités en utilisant Shibboleth.

Formats de métadonnées

ORI-OAI permet de gérer tous types de formats de métadonnées. Actuellement, le module de workflow propose un éditeur LOM et un éditeur Dublin Core. Pour les ressources pédagogiques, l'éditeur LOMfr sera proposé dès que le schéma XML LOMfr aura été finalisé dans le cadre des travaux de l'AFNOR. D'autres formats sont en cours d'intégration : TEF, CDM, AO.fr... D'ores et déjà, tout format de métadonnées exposés (CDM,...) peut-être moissonné et exploité au travers du moteur de recherche.

Instances d'ORI-OAI en production pour UNIT

Le portail UNIT (www.unit.eu), a été la première application opérationnelle du logiciel ORI-OAI. Il exploite les fonctions de moissonnage et d'interrogation.

La version 1.0 du logiciel ORI-OAI a été publiée le 17 Octobre 2007, elle est accompagnée d'un site web décrivant complètement la structure du logiciel, les procédures d'installation et d'utilisation.

Un logiciel de démonstration sous Windows est disponible et installable automatiquement, a fin mars, ce logiciel a été téléchargé 234 fois. Le logiciel de production a été installé dans les établissements suivants : Université de Bordeaux1, Insa-lyon, Université Paris10, UVHC, Université de Bretagne Sud, Université de Haute Bretagne, Université Toulouse2, Iufm de

Bretagne, Université Nancy2, Université de la Réunion, UPMF, Université Rennes 1, INP-INSEEIHT Toulouse, Université Paris-Sud, Université de Brest, IUFM de Bretagne, EDUCAGRI, Ecole de St-Cyr, Université du Mans.

Documentation du projet

Le site web : <http://www.ori-oai.org>

La description générale : http://www.ori-oai.org/vue_generale.html

Une présentation fonctionnelle : <http://www.ori-oai.org/node/47.html>

L'architecture détaillée : <http://www.ori-oai.org/architecture.html>

Les téléchargements et l'accès à toutes les documentations : <http://www.ori-oai.org/telechargements.html>

A fin mars, le logiciel de démonstration sous Windows a été téléchargé 234 fois. Le logiciel de production a été installé dans les établissements suivants : Université de Bordeaux1, Insa-lyon, Université Paris10, UVHC, Université de Bretagne Sud, Université de Haute Bretagne, Université Toulouse2, Iufm de Bretagne, Université Nancy2, Université de la Réunion, UPMF, Université Rennes 1, INP-INSEEIHT Toulouse, Université Paris-Sud, Université de Brest, IUFM de Bretagne, EDUCAGRI, Ecole de St-Cyr, Université du Mans.

Utilisation du logiciel

Le logiciel est utilisé comme outil de base pour la gestion du portail d'UNIT, depuis près d'un an, il n'est cependant pas limité à la manipulation de contenus pédagogiques, l'INSA de Lyon a pour objectif de gérer son archive institutionnelle, un développement a été financé par le ministère pour permettre le dépôt de thèse selon la norme TEF et un second pour se connecter à HAL (Hyper Articles en Ligne), le logiciel de dépôt d'archives ouvertes du CNRS.

Au cours de la démonstration, on montrera la moisson d'archives multiples de la BNF à HAL, en passant par le CNRS.