

ORI-OAI à l'Université Bordeaux 1

La diffusion des thèses



* www.u-bordeaux1.fr



UNIVERSITÉ BORDEAUX 1 | SCIENCES TECHNOLOGIES

Les Sciences et les Technologies au service de l'Homme et de l'Environnement



↘ Sommaire

- *Contexte et historique*

- *Le choix d'ORI-OAI*

- *Implémentation*

- *Perspectives et évolutions*

Contexte et historique (1)

▪ Numérisation et diffusion des thèses : une volonté forte du SCD depuis 2003.

× Cyberdocs, tamino xml server, sparte :le temps des expériences ...

‣ Epoque « tout xml » : tests sur la chaîne de conversion cyberdocs et intégration dans tamino XML server (base de données XML native avec APIs et outils de gestion). Nécessité d'un texte structuré (feuilles de style).

‣ Diffusion avec la plateforme SDX : recherches fines (richesse du xml) et compatibilité OAI-PMH

‣ SPARTE : projet de l'ABES et du CINES. Outil de conversion / diffusion en ligne pour les thèses. Démarré en 2004, il ne dépassera pas vraiment le stade de l'expérimentation pour s'arrêter définitivement en 2006.

‣ Premiers bilans : pas très positifs

‣ Lenteur du traitement, ressources humaines importantes, technicité

‣ problématique de la feuille de style, LaTeX non géré ...

x

Contexte et historique (2)

▪ La diffusion des thèses par la scolarité

× Le serveur de diffusion des thèses de l'Université

- Modernisation de l'imprimerie de l'Université : génération de la version imprimée de soutenance et d'une version de diffusion à partir d'un fichier remis par le doctorant.
- Gestion de la diffusion par la Scolarité (DISVE) : traitement et mise en ligne par un informaticien. Les thèses sont au format PDF. Une base annexe permet de décrire chaque thèse. Un petit moteur de recherche interroge les champs de cette base.
- Développement local qui a au moins le mérite d'offrir une visibilité des thèses et une recherche basique.

➢ Les manques ...

- Sur la base du volontariat : pas d'exhaustivité.
- Manque de visibilité hors de l'établissement
- × Pas de normes pour la description des thèses, recherche vraiment basique

Contexte et historique (3)

▪ Les publications scientifiques

- × Mise en place d'une maquette avec Eprints (2007)
 - Le SCD missionné par le VP du Conseil Scientifique pour mener une étude sur la possibilité d'une archive institutionnelle pour l'établissement.
 - Choix du logiciel EPRINTS (<http://www.eprints.org>) développé par l'Université de Southampton (Open Source).
 - Possibilité de traiter plusieurs types de documents (thèse, ressources pédagogiques, publications scientifiques diverses)
 - Compatibilité OAI-PMH
 - Interfaces conviviales et de nombreux outils sympathiques pour le déposant (templates, fils rss, exports pour bibliographies, gestion de versions, sherpa ...).
 - Workflow assez simple (un déposant, une validation « scientifique »)
- Mais ...
 - Problématique de HAL.
 - × Nécessité de développement pour intégration au SI de l'établissement. (ENT)

Contexte et historique (4)

▪ **Bordeaux 1 choisit le dépôt électronique des thèses**

× A compter du 1 septembre 2008 ...

- ›Après avis du Conseil Scientifique et décision du Président.
- ›Le SCD est chargé de la mise en place du dépôt électronique des thèses.
- ›Nécessité (arrêté du 7 août 2006) d'utiliser la plateforme STAR de l'abes.
- ›La diffusion des thèses via l'imprimerie et la scolarité n'a plus lieu d'être : il faut mettre en place de nouvelles procédures et un nouvel outil de diffusion (eprints ?)
- ›Circuit administratif complexe à Bordeaux 1 : présence d'outils locaux, d'outils plus institutionnels (Apogée thèse) auquel on ajoute encore une couche avec STAR et éventuellement une autre avec l'outil de diffusion.
Double ou triple saisies en perspective en attendant que ces systèmes soient capables de communiquer entre eux.
Peu satisfaisant pour l'ensemble des acteurs !

Le choix d'ORI-OAI

× **Le seul possible ?**

- Découverte du logiciel aux JRES 2007 (Novembre)
- Les points forts :
 - Projet Open Source avec des soutiens institutionnels forts (MENESR, SDBIS, SDTICE ...)
 - Projet ambitieux : un référentiel unique de documents connecté au SI de l'établissement et communiquant avec les plateformes nationales.
 - Indexation professionnelle
 - Utilisation de classifications et nomenclatures paramétrables et s'appuyant sur des standards (Dewey, etc ...).
 - Système de recherche avancée
 - Projet très lié à la communauté ESUP et donc s'intégrant complètement dans nos systèmes (systèmes d'authentification, annuaires, ENT)
- Parallèlement, le projet du Club EEA se met en place.

Implémentation d'ORI-OAI (1)

× Maquette

- Groupe de travail : AMIE, SCD, DRIMM. Travail confié au SCD. Réunions régulières sur la viabilité du projet.
- Installation et paramétrage : le temps de la prise en main technique.
 - 7 modules indépendants mais qui communiquent entre eux : architecture à bien comprendre.
 - Le Workflow : possibilités de paramétrages très riches et très fins. S'appuie sur l'annuaire LDAP de l'établissement pour récupérer un certain nombre de données. Authentification sur le SSO de l'établissement.
 - Le module Vocabulary : concept très intéressant notamment dans le cadre de la construction de nomenclatures communes (a minima au niveau de l'établissement mais aussi au niveau national).
 - Le module de recherche : configuration très riche. Choix des champs sur lesquels portent les recherches, création de recherche prédéfinies (par type de document, par domaine, par date, etc ...).
 - L'entrepôt OAI : paramétrage de sets

Implémentation d'ORI-OAI (2)

× **Maquette**

- Choix de s'appuyer sur la Dublin Core pour la description des thèses en attendant la prise en charge de TEF.
- Alimentation via le workflow de quelques thèses.

➤

× **Premiers bilans**

- Projet jeune (mi-2008 ...), utilisable pour la diffusion (en s'appuyant sur le module de recherche) mais pas encore pour la caractérisation de ressources numériques (dépôt impossible, lenteurs difficilement tolérables).
- Le temps d'appropriation n'est pas négligeable. Le point positif est que les technologies employées sont maintenant bien connues des services informatiques... (notamment pour les établissements qui ont déployé un ENT sur ESUP)
- Les possibilités entrevues lors du maquettage ainsi que les évolutions (indispensables) annoncées en font un projet réellement prometteur.

Implémentation d'ORI-OAI (3)

× Reprise de l'existant

- Toutes les thèses sur le serveur de diffusion ont été intégrées dans ORI-OAI
- Le manque de norme (champs textes libres, minuscules/majuscules, séparateurs au choix ...) a rendu le travail très fastidieux car tout n'a pas pu être automatisé. Problématique de l'encodage (UTF-8 de bout en bout)
- Reprise des thèses déposées dans TEL.
- Au final, plus de 500 thèses ont été intégrées.

× ORI-OAI dans le circuit de la thèse...

- Une fois la thèse validée dans STAR, on procède à un export des métadonnées TEF dans un fichier XML (fonctionnalité de STAR). Ce fichier est ensuite placé dans un répertoire avec le fichier PDF de la thèse.
- Un script transforme le fichier TEF en fichier DC, le zippe puis copie le fichier de la thèse vers le serveur de diffusion. Les métadonnées TEF sont conservées pour reprise ultérieure
- Le fichier zippé des métadonnées est ensuite importé dans ORI-OAI.
- Traitement par lots. Quelques secondes de traitement.

Implémentation d'ORI-OAI (4)

× Paramétrages du module Search

- Attention particulière portée à ce module qui est la partie visible pour les utilisateurs
- Choix d'une présentation mettant en valeur les nouveautés, une classification thématique par discipline (vocabulaire local basé sur les spécialités de la thèse et les écoles doctorales).
- Module de recherche avancée sur presque toutes les métadonnées.
- Création d'une recherche prédéfinie pour mettre en valeur les HDR.

×

× Démonstration ...

×

× Impacts :

- × Visibilité beaucoup plus importante (15 000 visites en 3mois). Indexation dans Google (et Google Scholar). Des doctorants se manifestent pour que l'on mette leur thèse en ligne.
- × Pas de gestion de droits : problème des thèses à diffusion différée ou confidentielle, traitement particulier à prévoir.

Perspectives et évolutions (1)

× Travaux à mener à court terme

‣ **Le module repository : l'entrepôt OAI.**

‣ Définition de sets

‣ Enregistrement de l'entrepôt dans différentes sources internationales de référencement.

‣ **Le module harvester : Moissonner des ressources distantes**

‣ Deux cibles prioritaires sont envisagées : HAL pour les publications scientifiques (limitations aux publications de l'établissement ?), les UNT pour les ressources pédagogiques.

‣ **Le module search : Intégration à l'ENT.**

‣ Pour l'instant le module de recherche est déployé sous forme de servlet et ne sert que pour une interrogation des thèses et HDR.

‣ L'idée est d'implémenter le module sous forme de portlet (directement intégré à l'ENT), en proposant des recherches profilées sur les thèses et HDR mais aussi sur les ressources moissonnées.

Perspectives et évolutions (2)

× Travaux à mener à moyen et long terme

‣ **Suivi des évolutions techniques :**

- Implémentation du module de stockage : esup-ecm (juin 2009 ?).
 - Possibilité de déposer un document
 - Gestion des droits
- Gestion des métadonnées TEF et DC+ pour les publications.

‣ **Mise en place de groupes de travail thématiques fonctionnels**

- Thèses : nécessité de redéfinir un circuit pour le dépôt électronique lorsque TEF sera implémenté. Avantage évident car la souplesse de paramétrage d'ORI-OAI permettra à coup sûr de mettre en place un workflow adapté aux spécificités de l'établissement (contrairement à STAR ...).
- Publications : travail déjà en cours avec une étude des usages existants et des ébauches des scénarii possibles. Dans tous les cas, il sera nécessaire d'avoir une alimentation automatique de HAL par ORI.

Perspectives et évolutions (3)

× Travaux à mener à moyen et long terme

‣ Les groupes de travail .

‣ Documents administratifs : pourraient être un levier « politique » pour le reste des documents. Beaucoup de choses à inventer mais un besoin réel.

‣ Documents pédagogiques : travail à mener avec un groupe technique. Des passerelles doivent exister entre la plateforme pédagogique (Claroline à venir pour Bordeaux 1) et ORI-OAI. Le projet CEPIA pourrait être une réponse.

‣ Les questions transverses

‣ Qu'est-ce qu'on moissonne (pertinence, dédoublement ..) ?

‣ Qu'est-ce qu'on exporte via l'entrepôt (problèmes de droits) ?

‣ Comment retrouve-t-on toutes ces ressources (paramétrage des recherches), qui les montre et les interroge ? (SID, Sites institutionnels, ENT)

‣ Le projet niveau PRES : Bordeaux 1 laboratoire ?

Conclusion :

- **ORI-OAI est un projet ambitieux, qui répond à un besoin réel des universités (gestion de la production numérique des établissements au sens large) mais qui nécessite un investissement important (adhésion politique indispensable).**
- **Le déploiement dans le cadre du SCD a permis de comprendre l'intérêt d'ORI-OAI. Cela a lancé le projet mais ça reste une implémentation mineure. L'intégration du module search dans l'ENT avec moissonnage de quelques ressources devrait être une prochaine étape importante.**
- **Nécessité d'un gros travail de réflexion niveau fonctionnel avant paramétrage technique (démarche inverse à Bordeaux 1 car contraintes de temps et problème de disponibilités des acteurs potentiellement impliqués)**
- **A ce jour, ORI-OAI reste un projet jeune avec de nombreuses évolutions à venir, indispensables pour une mise en production à l'échelle de l'établissement.**
- **En attendant l'implémentation des fonctionnalités basiques indispensables (a minima le dépôt), un établissement peut (doit ?) dès à présent étudier les divers aspects fonctionnels liés à la mise en place d'une GED. (démarche projet souhaitable).**
- **A noter la remarquable disponibilité des développeurs d'ORI-OAI pour un soutien technique et le gros travail de défrichage fait par les groupes fonctionnels (de nouvelles recrues seraient les bienvenues ...).**

QUESTIONS ?...



* www.u-bordeaux1.fr



UNIVERSITÉ BORDEAUX 1 | SCIENCES TECHNOLOGIES

Les Sciences et les Technologies au service de l'Homme et de l'Environnement

