



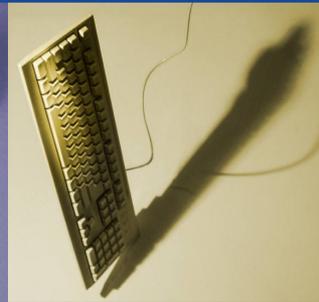
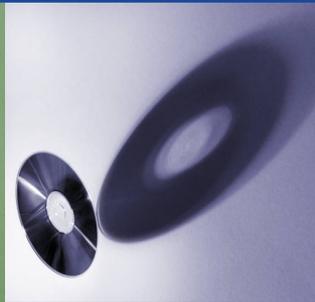
Projet ORI-OAI

Outil de **R**éférencement et d'**I**ndexation

Réseau de portails **OAI**

Université Ibn Zohr

Agadir, 12 mai 2008



Introduction

Contour fonctionnel

Concepts

Mise en oeuvre

Évolutions et projets connexes

Bilan



Introduction

- Les établissements universitaires
 - Croissance de la création numérique institutionnelle (scientifiques, pédagogiques, administratifs, documentaires)
 - Accessibilité aux ressources dans divers contextes
 - Référentiel unique de documents, indexation professionnelle
 - «Système Global d'Information»
 - La structuration initiée par les projets nationaux (UNT, UNR, SI d'établissement)

- Les UNT (Universités Numériques Thématiques)
 - Réflexion UNIT : évolution du portail et choix d'une organisation en « réseau de portails OAI-PMH»
 - Convergence sur les besoins de référencement et d'indexation avec l'UNT UVED
 - Nécessité de mutualiser les outils et les moyens d'accompagnement pour favoriser la démarche avec l'ensemble des UNT
 - ORI-OAI : projet Open Source inter-UNT soutenu par la SD/TICE

- Un groupe projet (établissements: Insa de Lyon, Université de Rennes 1, INPT, Université de Valenciennes)

- Un pilotage inter-UNT

- Une démarche d'ouverture et d'échange
 - Site Web et listes de diffusion
 - Groupes de travail (ressources pédagogiques, thèses, archives ouvertes)

Établissements



UNT



Avec le soutien du MENESR





Projet ORI-OAI

Insertion de ORI-OAI dans l'établissement



- Pourquoi intégrer ORI-OAI dans un établissement?
La réponse est fonction de l'utilisation que l'on a de l'outil :
 - ORI-OAI comme portail de ressources numériques
 - Diffusion et partage des ressources
 - Publication des productions de l'établissement
 - ORI-OAI comme archive institutionnelle
 - Diffusion et partage des ressources
 - Publication des productions de l'établissement
 - Conservation du patrimoine documentaire

■ Mise en place de ORI-OAI dans l'établissement

Deux modes de mise en place qui dépendent du contexte de l'établissement :

- ORI-OAI doit répondre a un besoin global
 - Une réflexion transversale à mener
 - Découpage du projet en plusieurs sous-projets
 - Prise en compte des spécificités de chaque sous-projet

- ORI-OAI doit répondre a un besoin précis
 - Action rapide inscrite dans une démarche à plus long terme
 - Peut constituer un test pour un projet de plus grande envergure

- Éléments d'une démarche de projet de gestion des documents numériques :
 - Une volonté politique en appui du projet
 - Une communication autour du projet
 - Au près des utilisateurs tout au long du projet
 - Au près de la communauté
 - Au près des instances internes concernées
 - Mise en place d'un groupe de travail transversal impliquant tout les acteurs
 - Bibliothécaires
 - Cellule TICE
 - Informaticiens

- Insertion de ORI-OAI dans le Système d'Information:
Principe de complémentarité avec les applications déjà existantes :
 - Utilisation des référentiels de l'établissement (LDAP...)
 - Système d'authentification / contrôle des accès (CAS, Shibboleth, ...)
 - Intégration à l'ENT
 - Interopérabilité avec l'ensemble des briques du SI (Moodle, GRAAL, HAL, application de la scolarite, ...)
 - Référentiel des documents numériques pour le SI (plateforme pédagogique, moteur documentaire, sites web de communication, ...)

Contour fonctionnel

- **Gérer et publier** les ressources numériques de l'établissement
 - **Référentiel unique** de ressources connecté au SI
 - **Production** de métadonnées et **diffusion** des ressources produites par l'établissement

- **Valoriser** la production de l'établissement par une indexation de qualité
 - **Indexation avec des normes et standards**
 - Utilisation de **classifications paramétrables** articulés sur des standards comme Dewey

- **Partager** les ressources numériques avec d'autres établissements
 - **Exposition** des fiches de métadonnées de l'établissement
 - **Agrégation** de fiches de métadonnées

- **Accéder** aux ressources numériques à **distance** en fonction des **droits d'accès**
 - Système de **recherche avancée** et **thématique**
 - Définition des **droits**

- Référentiels partagés
- Système d'identification et d'authentification de l'établissement
- Intégration dans l'ENT
- ...

recherche	sur les métadonnées, texte intégral
diffusion	en respectant le droit d'auteur, les licences et les droits d'accès
publication web	versions de publication, workflows, plusieurs schémas de métadonnées
archivage	documents natifs toujours disponibles pour les auteurs
partage	protocole OAI/PMH, fédération d'identités
référencement	indexation de qualité adaptée aux divers types de documents, classification pivot

- **Ressources d'enseignement**
 - ressources pédagogiques multimédia, QCM, exercices, photocopiés numériques, ...

- **Travaux d'étudiants**
 - rapports de stage
 - synthèses bibliographiques

- **Travaux de recherche**
 - publications, pré-publications, rapports techniques
 - Thèses, mémoires de Master

- **Ressources documentaires éditoriales**
 - revues, livres électroniques

- Groupes de travail par types de ressources
 - Ressources pédagogiques numériques
 - Thèses
 - Archives ouvertes
 - Documents administratifs ?
 - ...

- Financement coopératif par les établissements

- Outil de **R**éférencement et d'**I**ndexation
 - mettre en réseau des ressources numériques
- + Protocole **OAI/PMH**
 - échanger des données
- + **.ORG**
 - développé en logiciel libre
 - pour les UNT, UNR, les établissements universitaires, ...
- = **ORI-OAI.ORG**



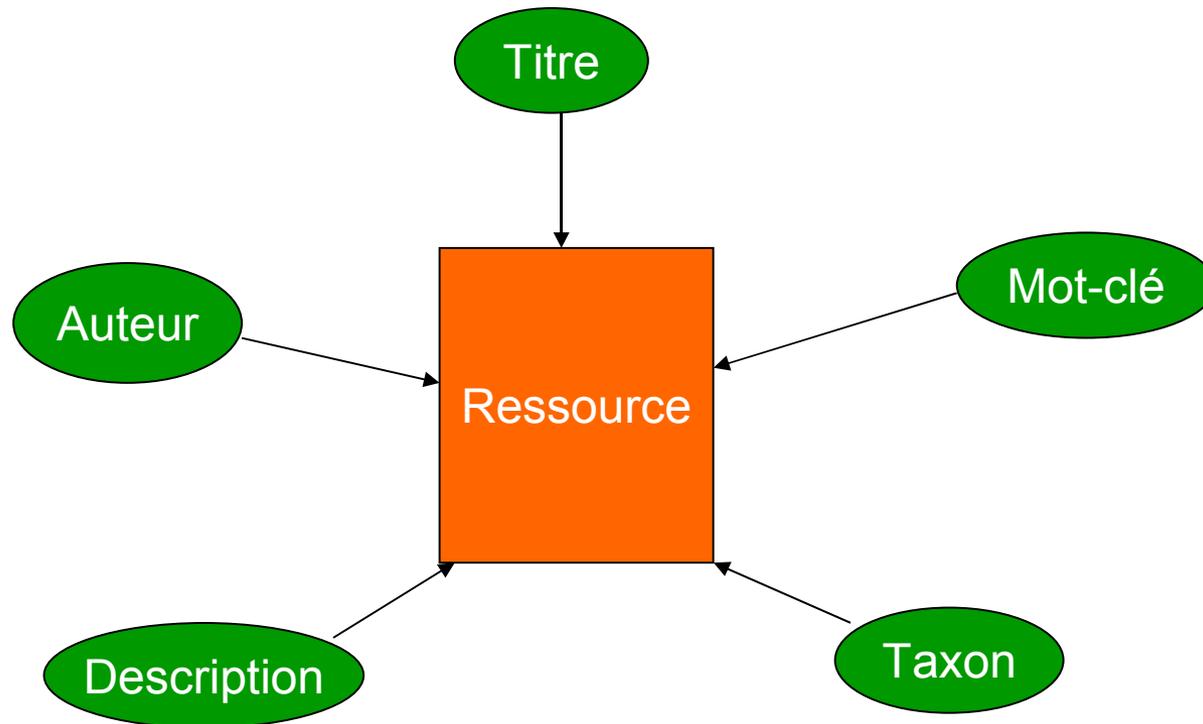
Concepts



Projet ORI-OAI

Concept de métadonnées

- Données liées à une ressource
 - Attributs communs à un ensemble de ressources



- Données liées à un contexte d'utilisation
 - Format générique (Dublin Core)
 - Format spécifique
 - Pédagogique : LOM, LOMFR, SupLOMFR
 - Offre de formation : CDM
 - Thèse : TEF
 - Bibliographique : MARC

- Créé en 1995 à Dublin (Ohio, USA) à l'initiative de l'OCLC
- Objectif : décrire les ressources numériques
- Définit une liste de 15 éléments « de base » permettant de décrire n'importe quelle ressource
- Possibilité de raffiner certains éléments (Dublin Core qualifié)
- Existence d'un schéma XML

■ Dublin Core simple

Ressource

15 éléments DC :

- Titre
- Sujet
- Description
- Date
- Langue
- Type
- Couverture
- Contributor
- Creator
- Publisher
- Format
- Identifier
- Source
- Relation
- Rights

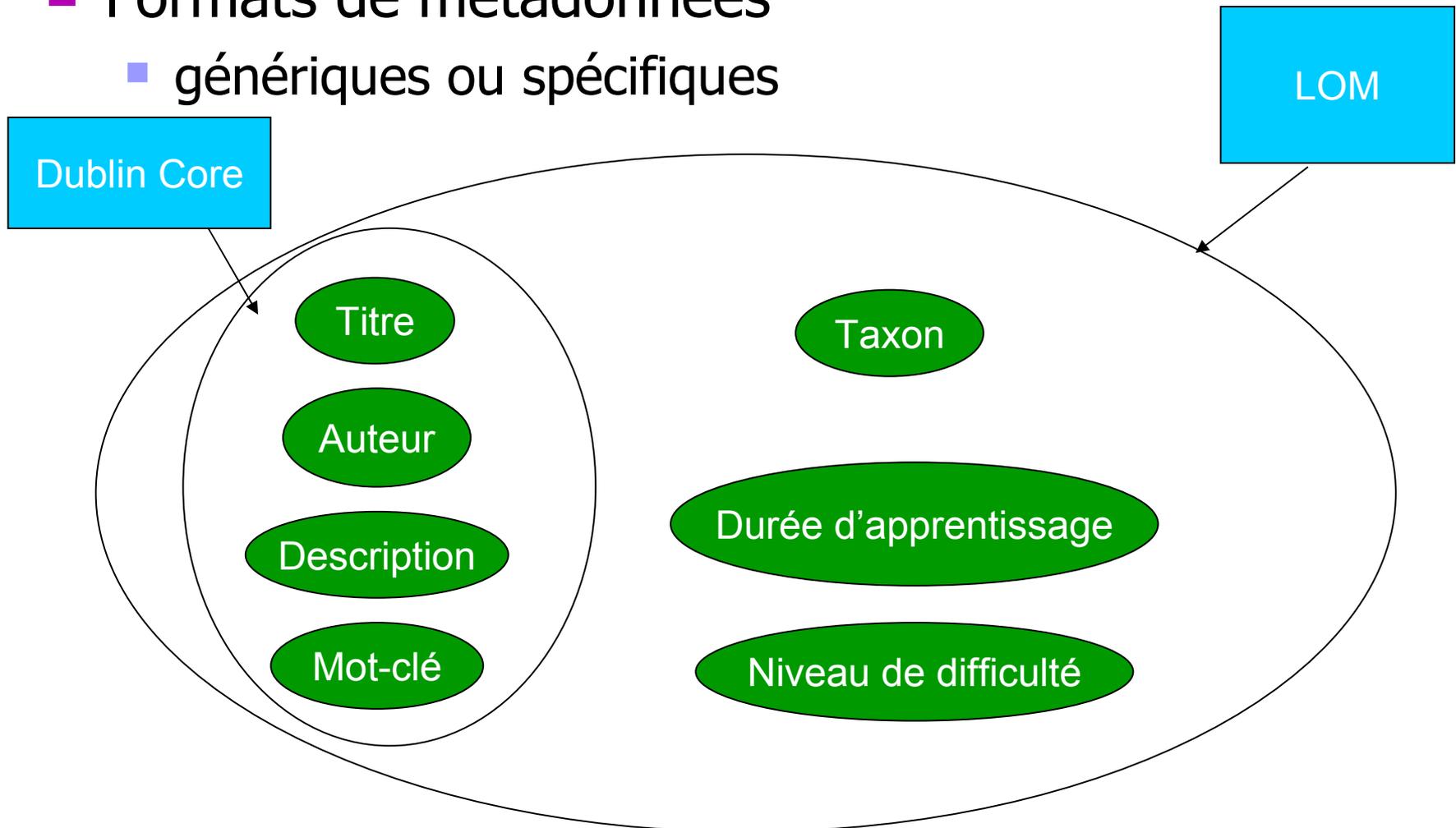
descriptives

agents

techniques

structurelles

- Formats de métadonnées
 - génériques ou spécifiques



- Format de métadonnées semi-structuré
 - basé sur un schéma XML
 - structure imbriquée, récursive
 - éléments répétables

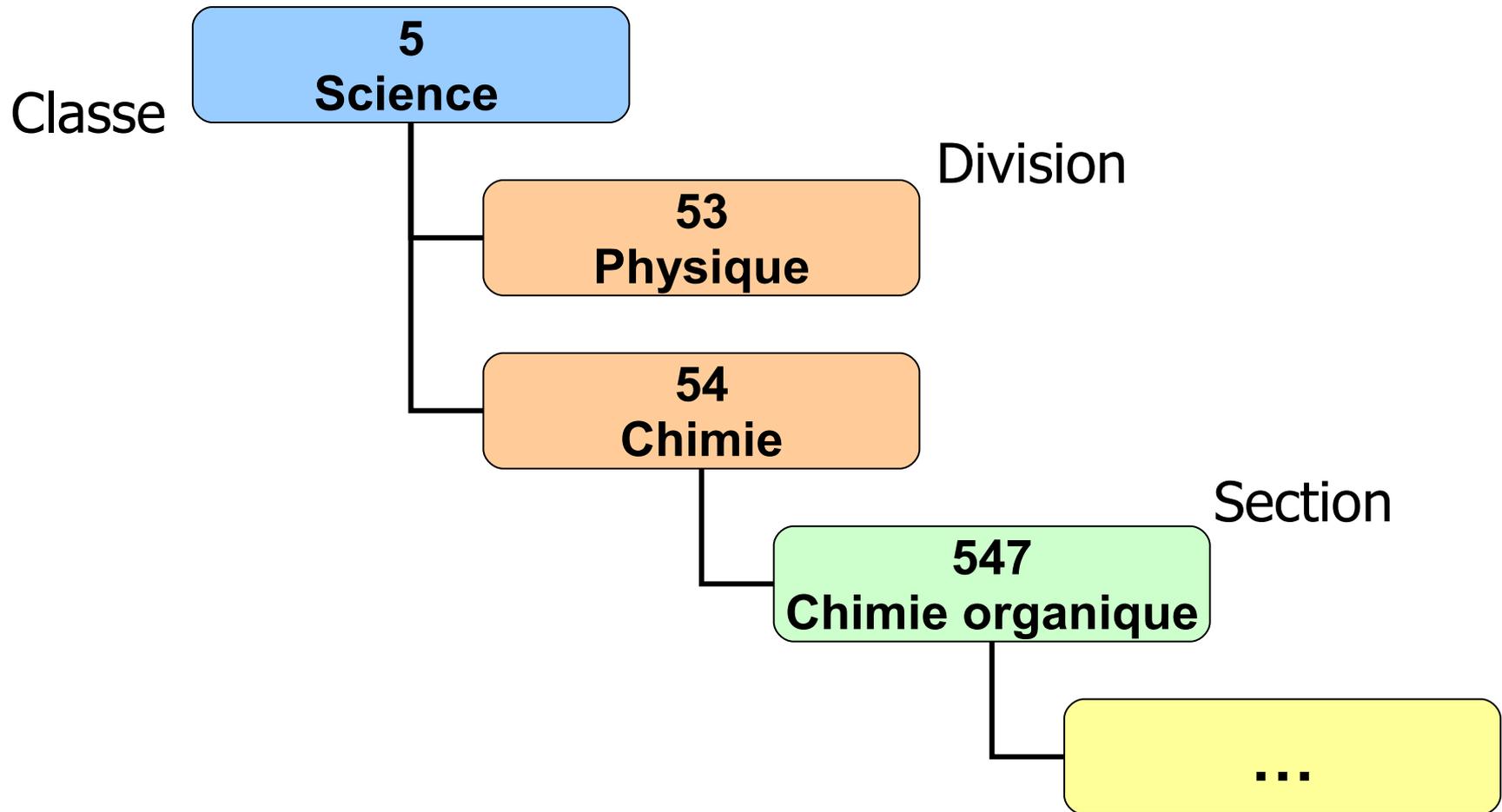
- Rôle des métadonnées dans un entrepôt de données
 - mise en commun d'attributs, constitution d'ensembles
 - qualité de référencement
 - vocabulaires fermés, saisie contrôlée
 - gestion de la cohérence des données et des redondances



Projet ORI-OAI

Classification

- Quoi ?
 - Langage documentaire
 - **Organisation des connaissances en domaines**
 - **Structure arborescente**: des notions générales aux détails
 - **Représentation** d'un domaine par un **indice numérique**

Un exemple :

- Application ?
 - **Classement physique** des documents dans la bibliothèque

- Comment ?
 - **Traduction** du contenu de documents en **indice(s)** (indexation systématique)

Distinction des indexations

systématique → accès thématique (par domaine, point de vue)

analytique → accès par sujet

- Objectif ?
 - construction d'un « **plan de classement** »
(autrement dit d'une classification simplifiée)

- Comment ?
 - définition de **domaines spécifiques** aux besoins de la bibliothèque
 - **caractérisation** de ces domaines, par des **indices Dewey** ou leur combinaison
 - **exploitation** de l'**indexation** Dewey des **documents**

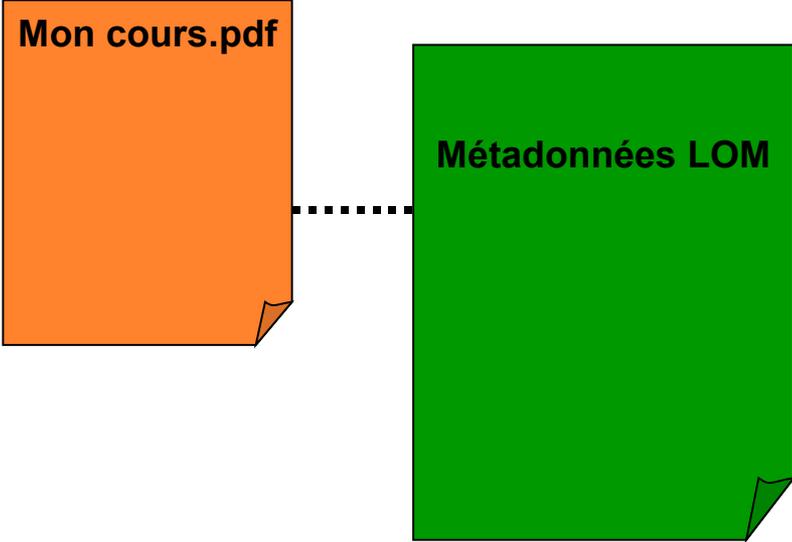
classification Dewey

=

pivot pour la constitution de **classifications spécifiques**

Approche **transposable** aux
ressources numériques

(avantages : "bibliothèque" virtuelle ; classement possible
d'une ressource dans plusieurs domaines ; réorganisation
rapide et souple ; ...)



Mon cours.pdf

Métadonnées LOM

Titre = Technologies énergétique

Classification Dewey = **621.45** (Aéromoteurs, éoliennes)

etc.

Classification UNIT

Energétique, énergie

Technologies de l'énergie éolienne

...

- 333.92*
- 621.312 136*
- 621.45*

Electricité et électrotechnique

Systèmes pour énergie renouvelable

...

- 621.312 136*
- 621.312 42*
- 621.312 44*
- 621.45*
- 629.229 5*

...

Classification Valenciennes

Sciences, techniques, sport

Génie mécanique et énergétique

• 621*

...

...



Projet ORI-OAI

Principe du référencement et de la diffusion

Doc

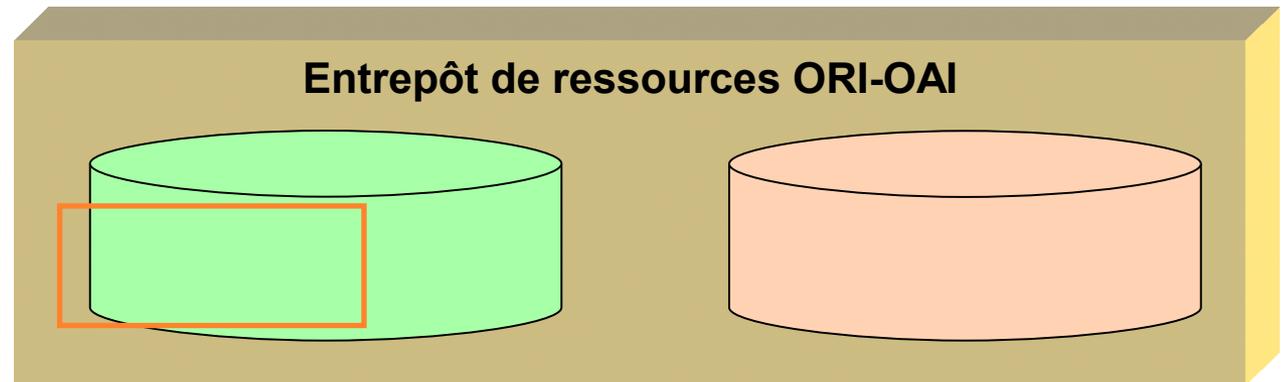


Auteur

Étape de saisie
auteur

Metadonnées

À publier





Bibliothécaire

Document

Titre

Description

Auteur

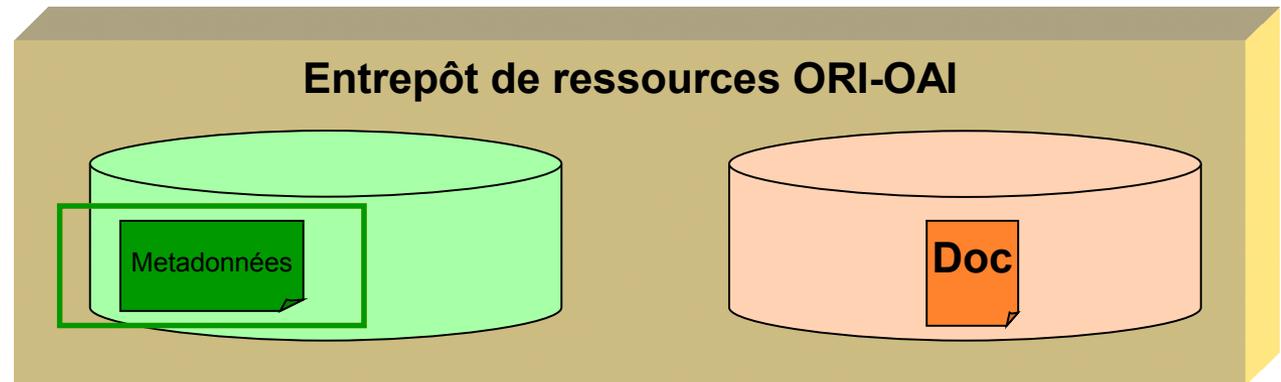
Classification

Publier **Rejeter**

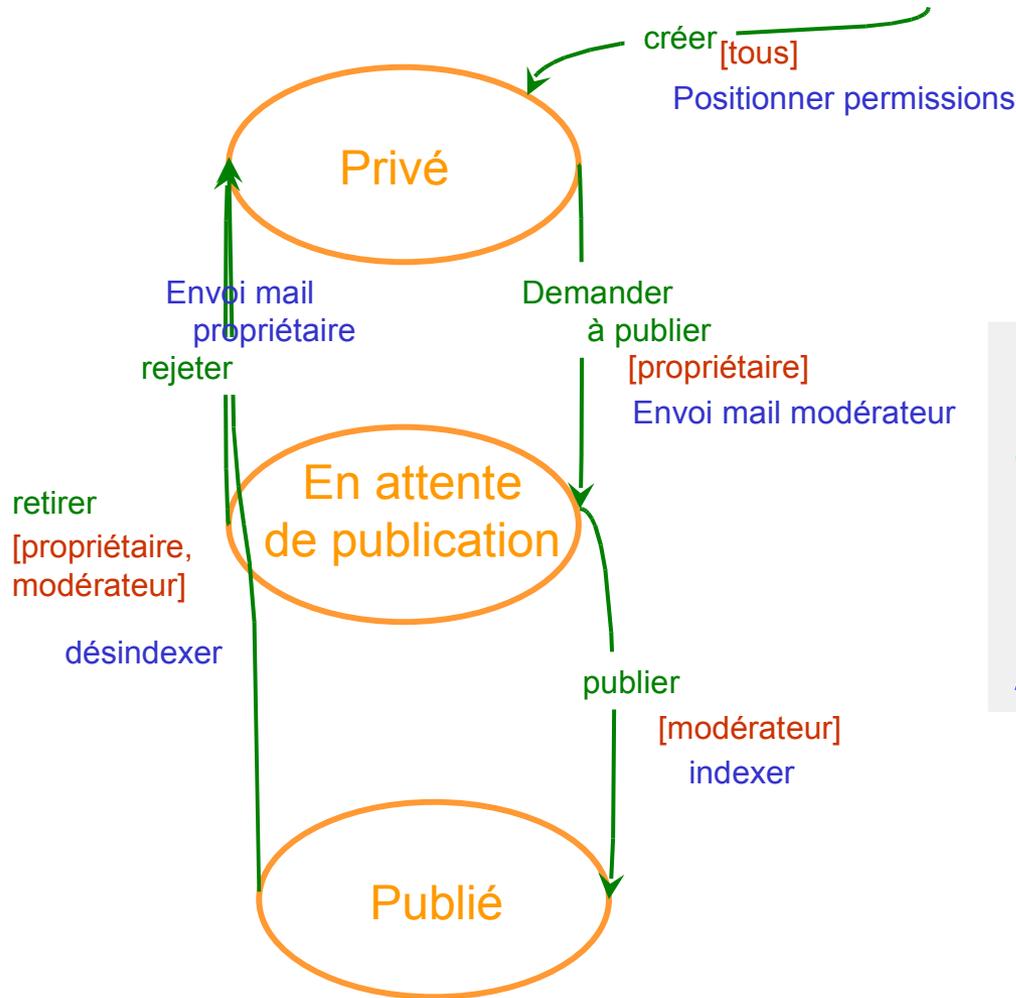
Étape de saisie
bibliothécaire

Metadonnées

Publié



Schématisation du processus : diagramme Etats / Transitions

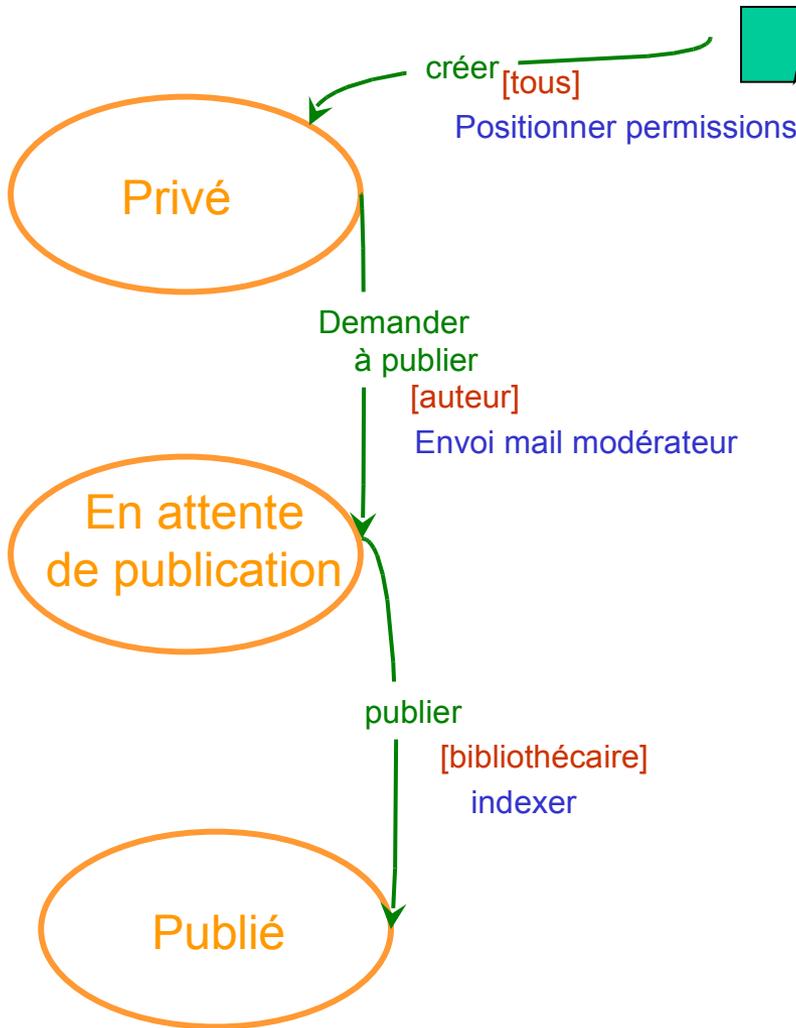


Etats (empruntables par la fiche)

Transitions (changements d'états)

Conditions (pour les transitions)

Actions (lors de transitions)



Déroulement ...
(demande de publication de la fiche acceptée)

Etats

Transitions

Conditions

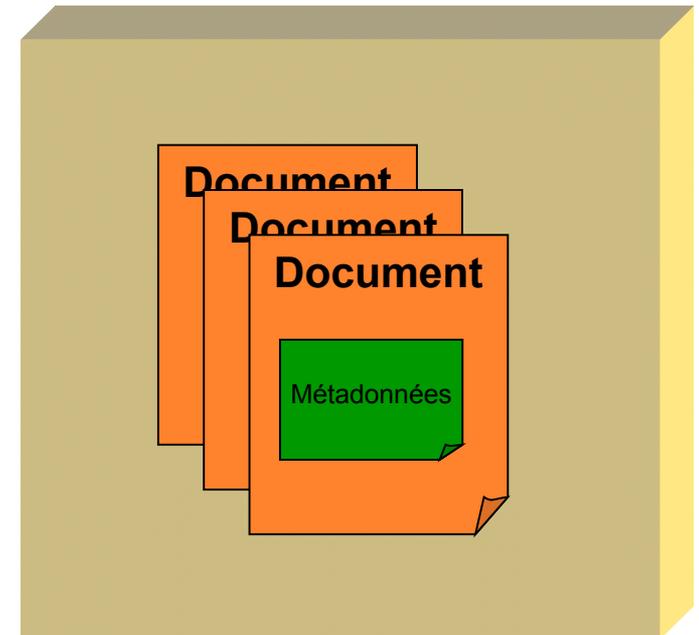
Actions



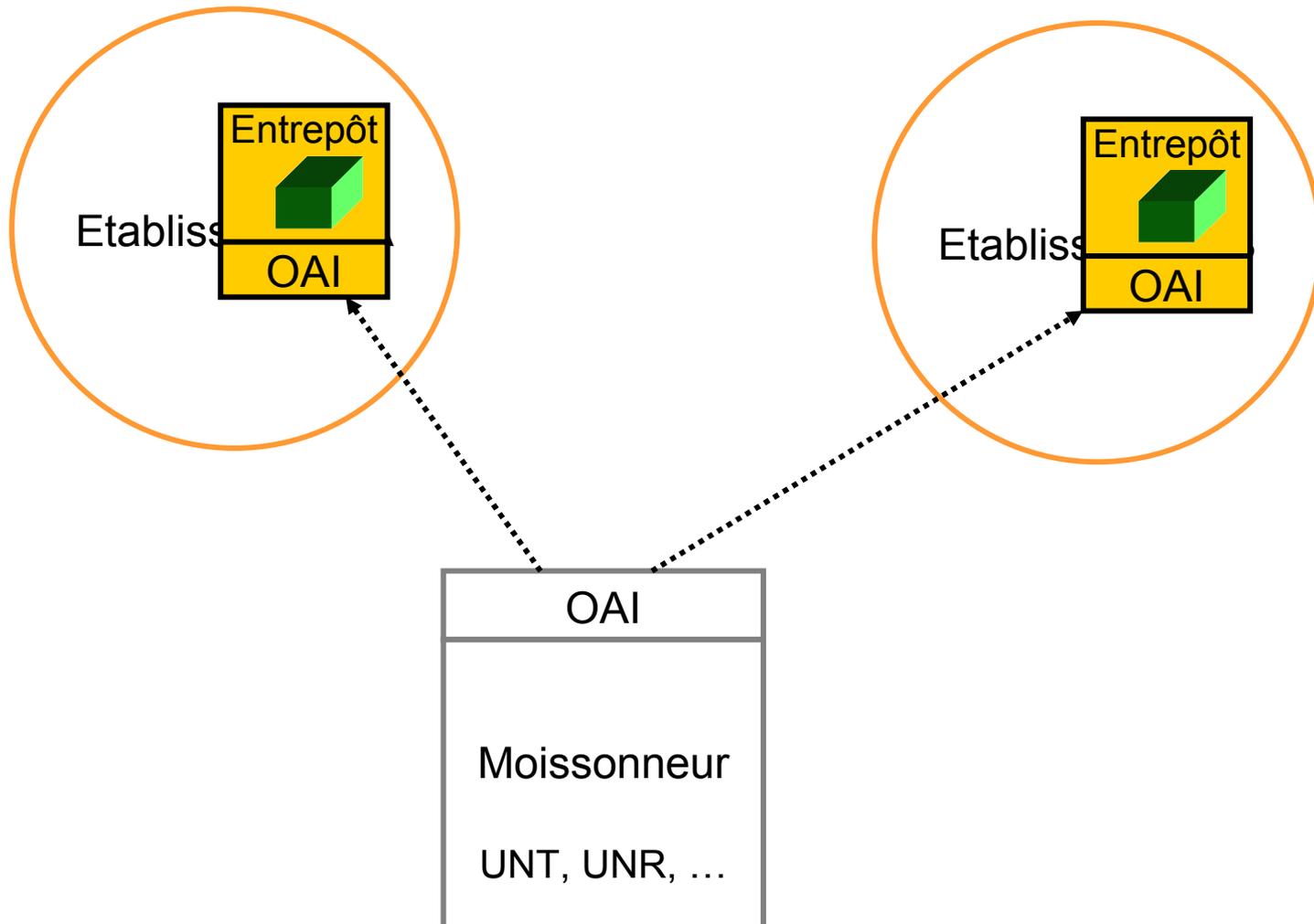
Projet ORI-OAI

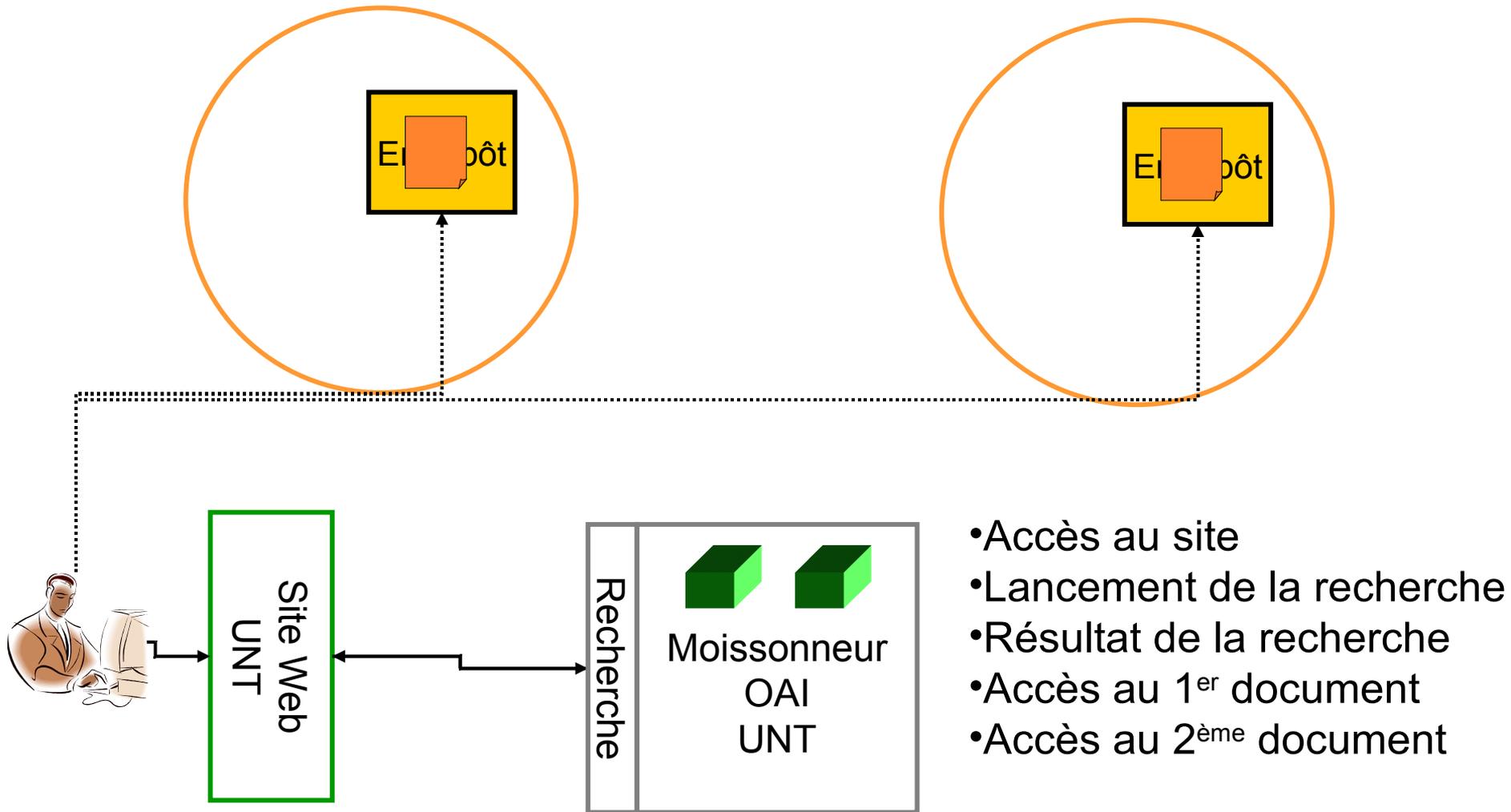
Principe des échanges

- Qu'est-ce qu'un entrepôt ORI-OAI ?
 - Stockage des ressources numériques
 - Métadonnées sur les documents
 - LOM
 - Dublin Core
 - Etc.

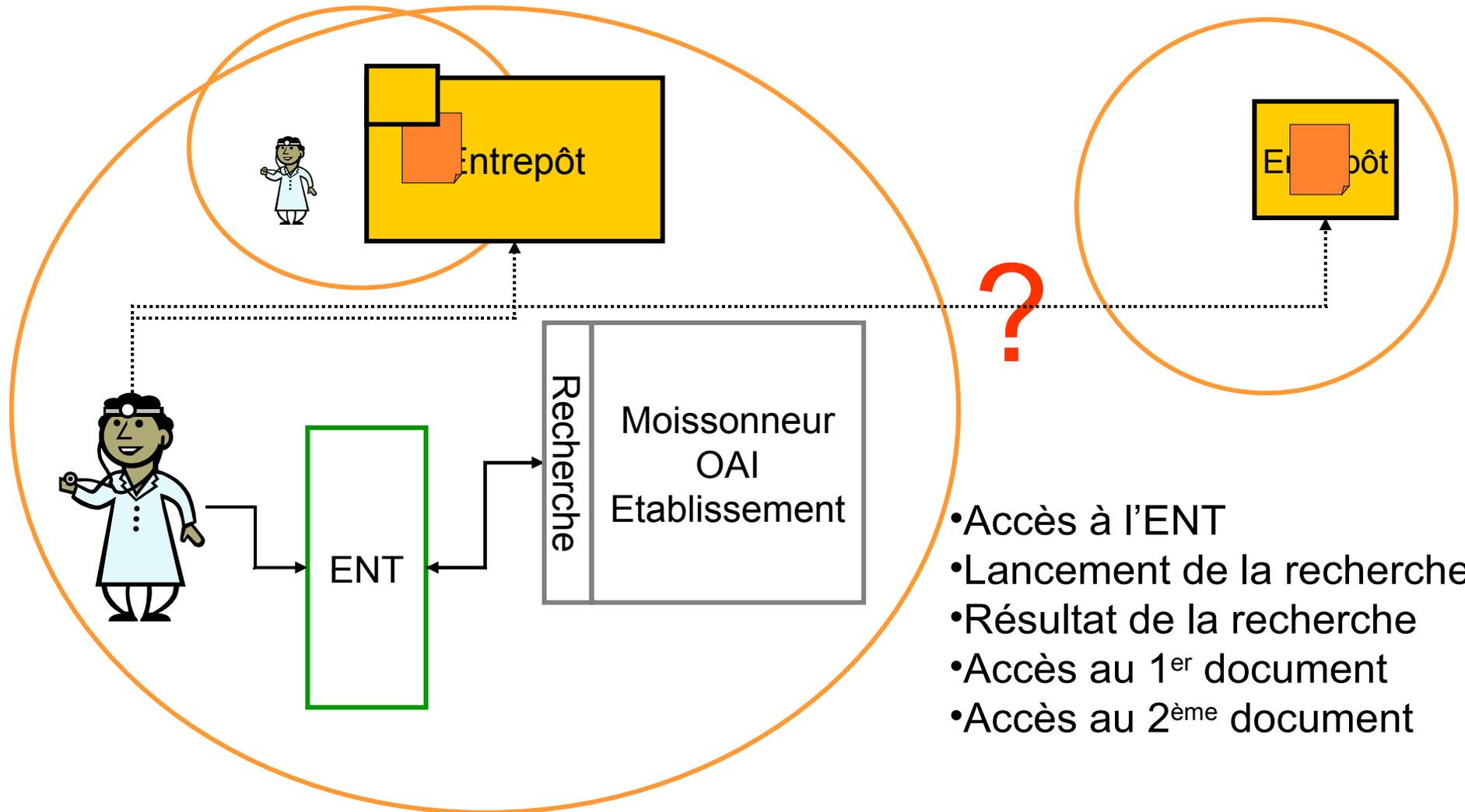


- Protocole informatique d'échange de données
- Permet la collecte et le partage de références
- Fait intervenir deux types d'acteurs:
 - Fournisseur de données (*data provider*) :
 - met en forme les métadonnées (au moins en DC)
 - les stocke dans un entrepôt (*repository*)
 - définit éventuellement des ensembles (*sets*)
 - Fournisseur de services (*service provider*) :
 - envoie des requêtes aux entrepôts (grâce à un moissonneur)
 - récupère les métadonnées exposées correspondantes
 - les exploite en fournissant des services à valeur ajoutée

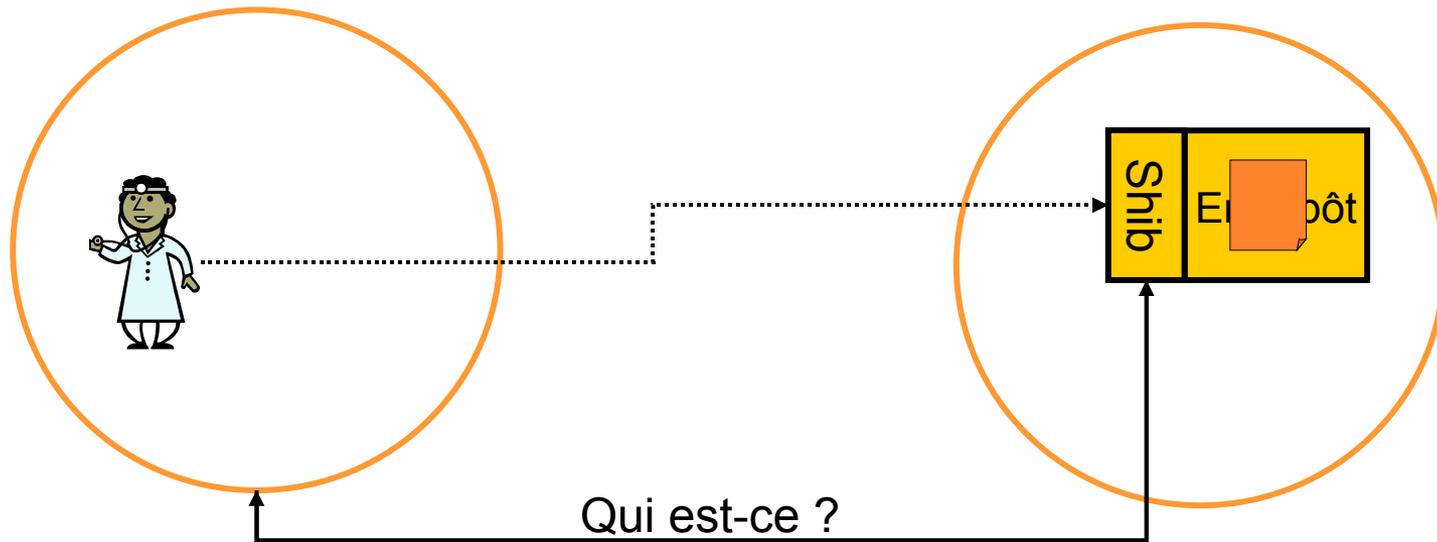




- Accès au site
- Lancement de la recherche
- Résultat de la recherche
- Accès au 1^{er} document
- Accès au 2^{ème} document

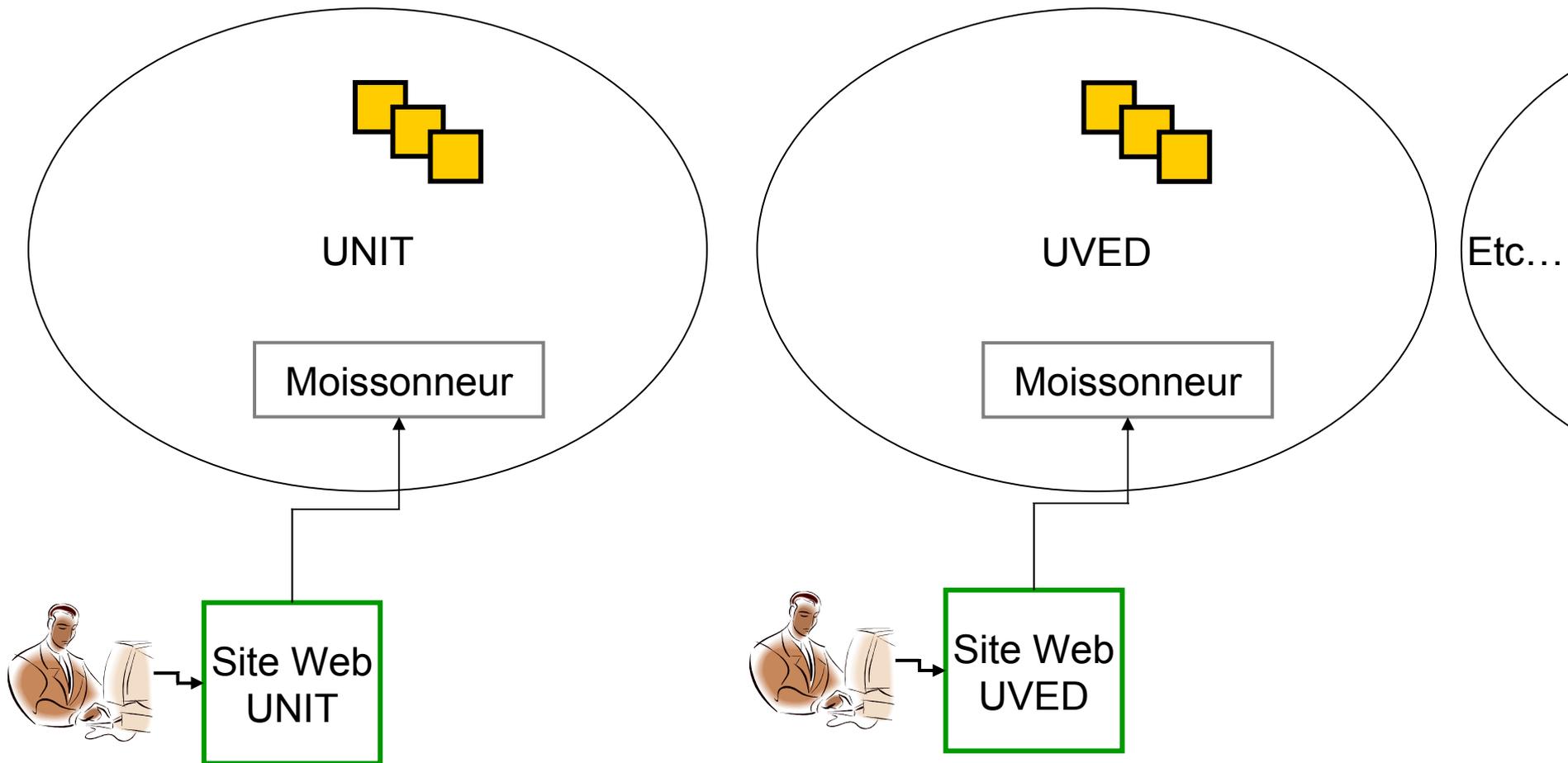


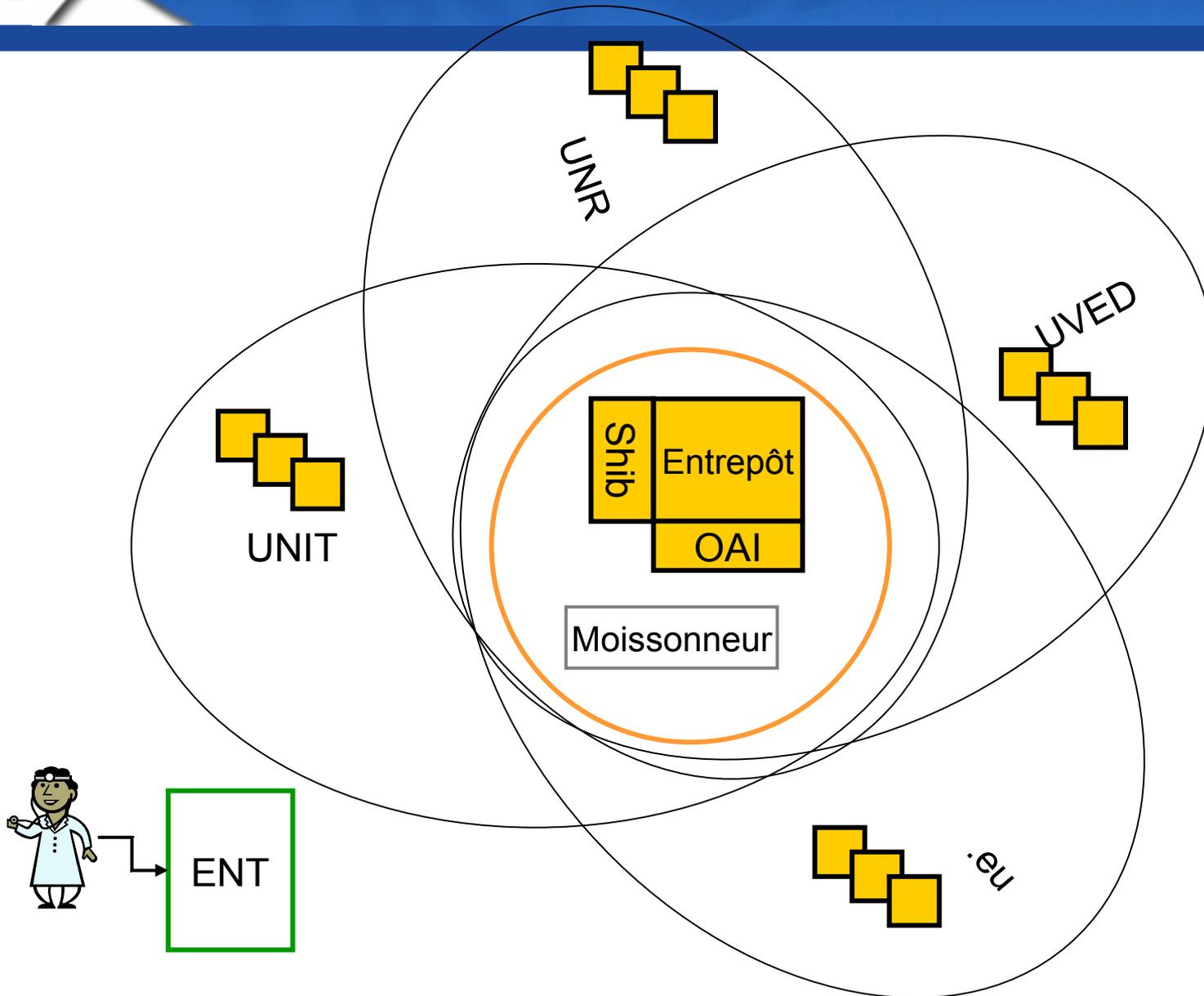
- Accès à l'ENT
- Lancement de la recherche
- Résultat de la recherche
- Accès au 1^{er} document
- Accès au 2^{ème} document



Un étudiant en médecine 2^{ème} année

- Tentative d'accès au document
- Requête à l'établissement « fournisseur d'identités »
- Réponse à l'établissement « fournisseur de services »
- Accès au document





Mise en oeuvre

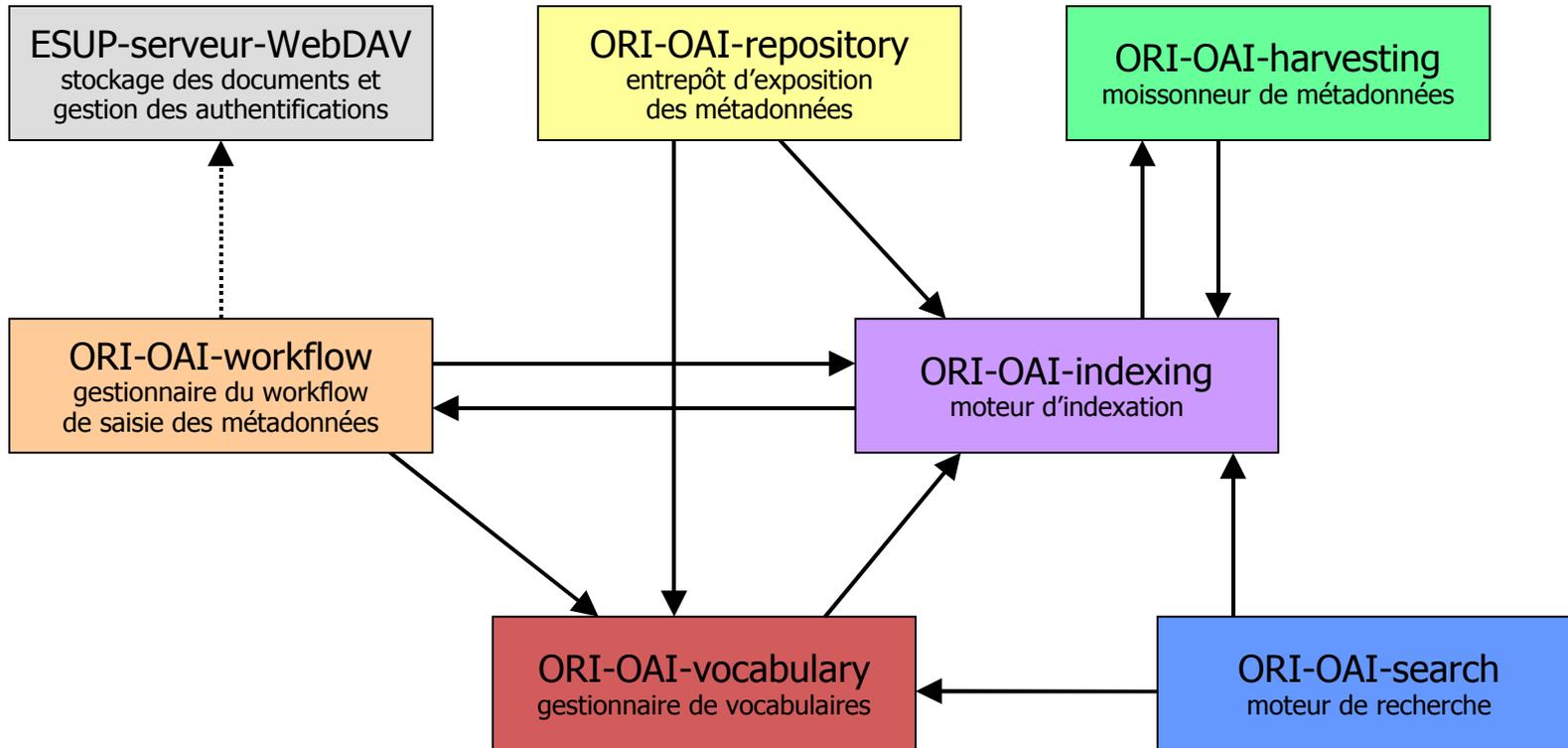


Projet ORI-OAI

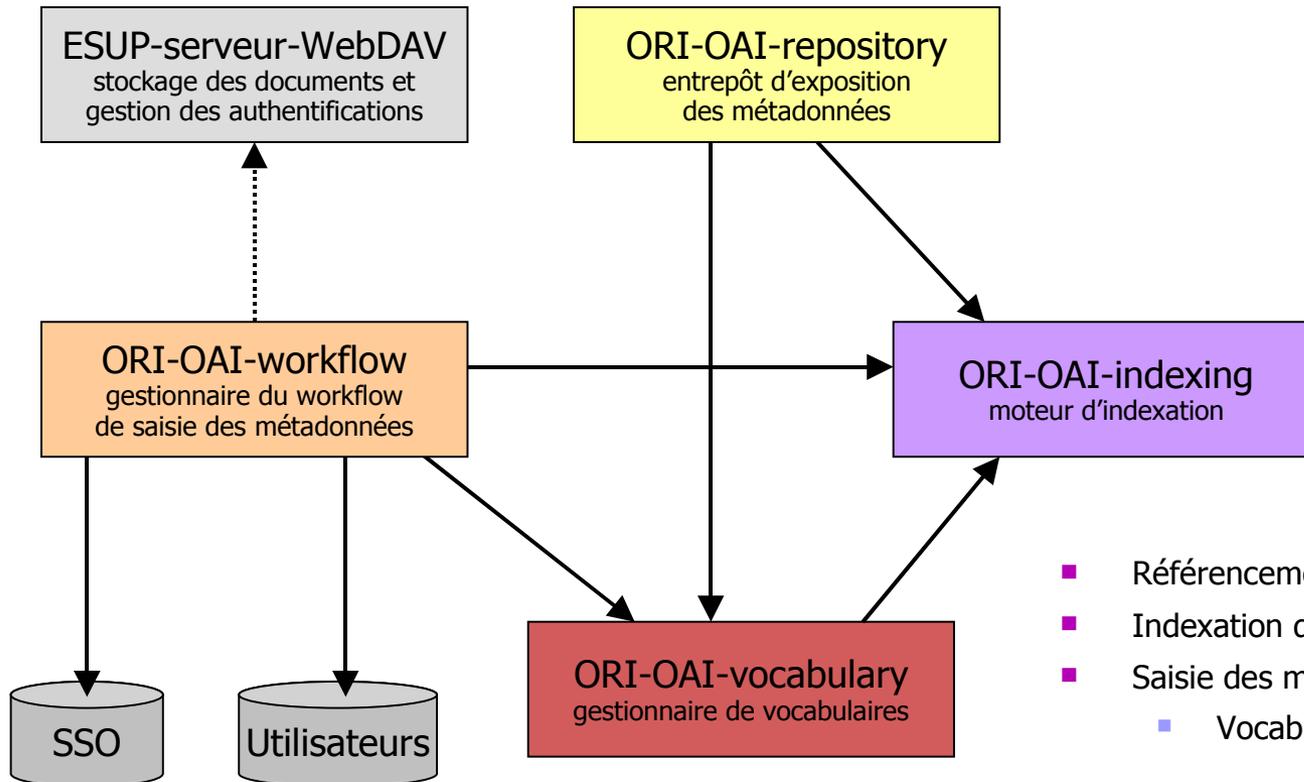
Architecture

- 7 modules
 - Indépendants
 - Ayant chacun un rôle précis
 - Tous interopérables
 - Dialoguent entre eux par Web Services

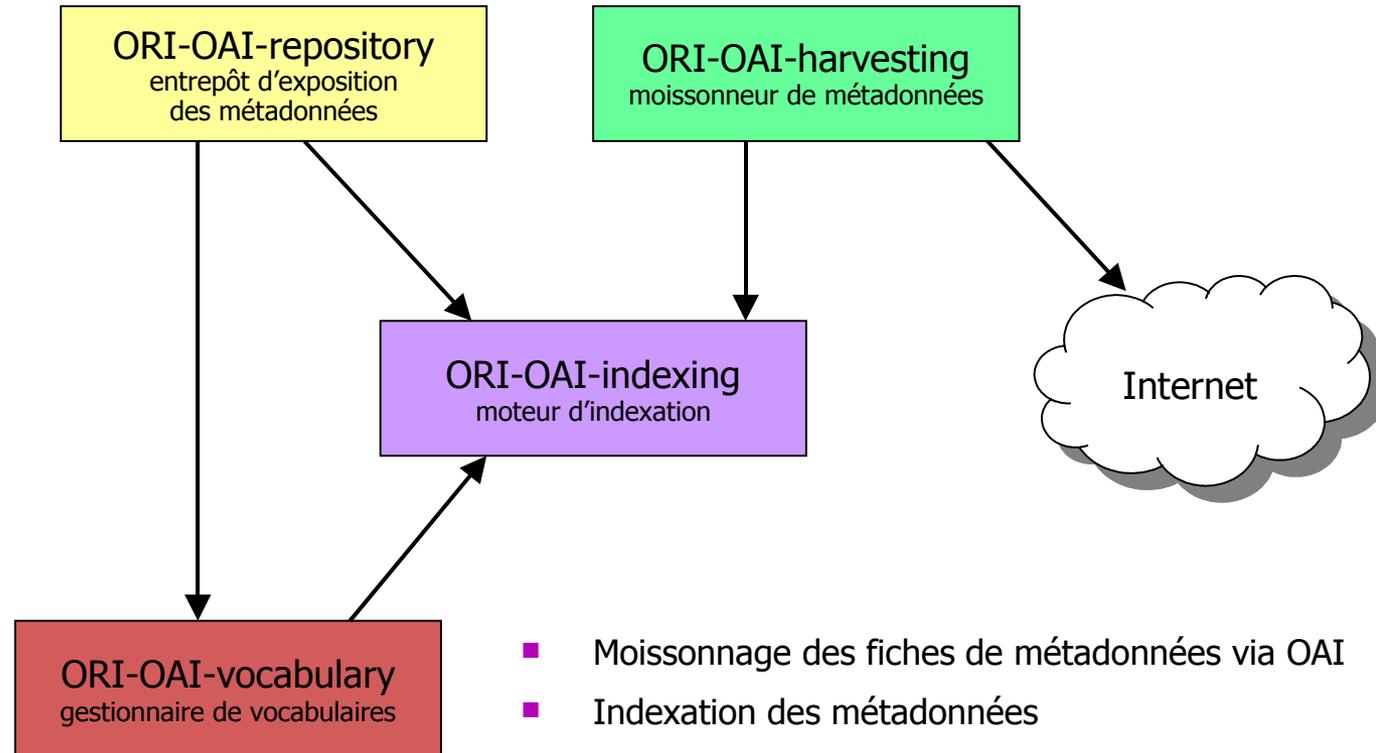
- Chaque module est très configurable
 - Plusieurs configurations sont proposées par défaut pour s'adapter au mieux aux besoins des différents établissements
 - Possibilité de faire évoluer ou d'intégrer d'autres configurations



- 7 modules
- interopérables



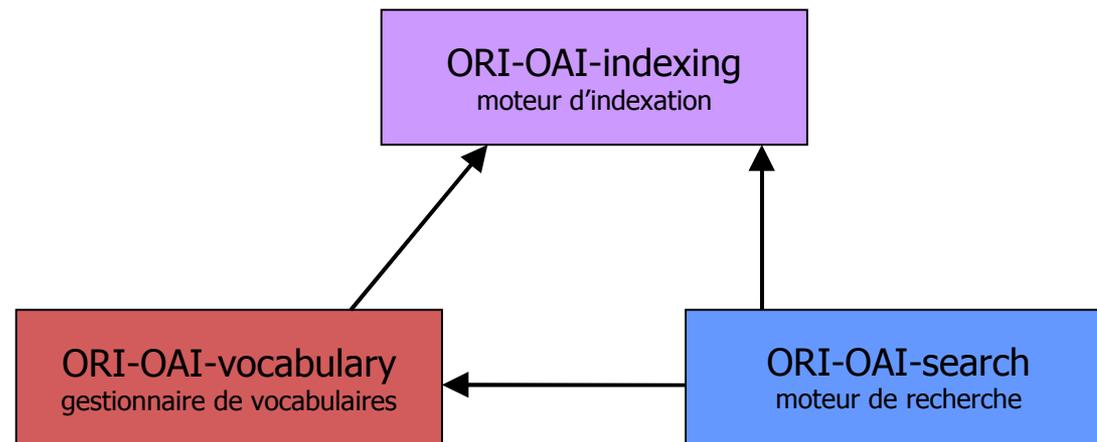
- Référencement du lien vers le(s) document(s)
- Indexation des métadonnées
- Saisie des métadonnées via vocabulaires
 - Vocabulaires dynamiques basés sur l'index
- Exposition OAI (optionnelle)
 - Exposition des fiches par requêtes sur l'index
 - Génération de sets OAI via le vocabulaire



- Moissonnage des fiches de métadonnées via OAI
- Indexation des métadonnées

- Exposition OAI (optionnelle)
 - Exposition des fiches par requêtes sur l'index
 - Génération de sets OAI via le vocabulaire

- Génération des interfaces à partir du vocabulaire
- Recherche dans l'index
- Traduction de certains résultats via vocabulaire





Projet ORI-OAI

Développement

- Développement Java
- Technologies et standards
 - Dublin Core
 - LOM / LOMFR
 - XML
 - Lucene/Lius
 - OAI-PMH
 - Shibboleth
 - Spring
 - JSR-168 (*intégration dans ENT, CMS, etc.*)
 - Etc.
- Licence GPL

- **Version 1.0.0**
 - Disponible depuis le 15 octobre 2007
 - Supporte différents types de documents
 - Édition de métadonnées
 - Ressources pédagogiques (LOM)
 - Ressources « génériques » (Dublin Core)
 - Moissonnage par OAI de ressources extérieures
 - Tout format de métadonnées
 - Pré-configuré pour Dublin Core, LOM et CDM
 - **Types d'installation**
 - De production
 - Configuration avancée
 - De démonstration
 - Auto-install d'une application Windows

- ORI-OAI version 1.0.0 est mis en place (ou en cours) par :
 - UNIT
 - UNR Nord-pas-de-Calais
 - Université de Rennes 1
 - Université de Valenciennes
 - INP de Toulouse / ENSEEIHT
 - Université de Nancy 2
 - Université de Paris 5
 - Université de Paris 10
 - Université de Paris Sud
 - Université de Bretagne Sud
 - Université de Limoges
 - INSA de Lyon
 - IUFM de Bretagne
 - Université de Brest
 - Université de Bordeaux 1
 - Université du Mans
 - Université de Versailles
 - St-Cyr
 - Educagri
 - Etc.

Évolutions et projets connexes

■ Version 1.5.0 (juin 2008)

■ Workflow

- Dépôt des documents depuis le workflow
- Gestion des droits d'accès aux documents depuis le dépôt
 - Utilisateurs locaux, groupes locaux et distants (shibboleth)
- Mise en place d'un protocole de communication avec d'autres plates-formes

■ Moteur d'indexation

- Indexation en texte intégral
- Crawler web pour les documents distants

■ Moteur de recherche

- Flux RSS
- Personnalisation des interfaces suivant le profil utilisateur
- Nouveaux modes de recherche

- Gestionnaire de vocabulaires
 - Gérer des vocabulaires distants et centralisés
- Système de stockage
 - Refonte du système
- Évolutions générales du projet
 - Identifiants et URL d'accès des ressources uniques et pérennes
 - Statistiques
 - De recherche
 - D'accès aux documents

- Version beta 1.6 (juillet 2008)
 - Référencement et publication dans l'archive de l'établissements
 - des thèses avec le format TEF
 - des publications de la recherche

- Version 2.0.0 (octobre 2008)
 - ORI-OAI / STAR
 - Export de métadonnées TEF vers l'archive nationale STAR

 - ORI-OAI / HAL
 - Échange des métadonnées et des ressources de l'archive de l'établissement avec l'archive nationale HAL

- CEPIA
 - Référencement dans ORI-OAI depuis une plate-forme LMS et un outil de chaîne éditoriale

Bilan

- Fruit d'un travail collaboratif efficace
 - Analyse des besoins multi-métiers
 - enseignants/chercheurs
 - cellule TICE
 - bibliothécaires
 - informaticiens
 - etc.
 - Rédaction des spécifications par les groupes de travail
 - inter-établissements
 - inter-UNT

- Paramétrage avancé pour répondre aux caractéristiques de chaque établissement

- Logiciel évolutif et libre

- Site du projet
 - <http://www.ori-oai.org>

- Contacts

Liste de discussion **ori-oai-utilisateurs**

<http://listes.univ-rennes1.fr/wws/info/ori-oai-utilisateurs>

Liste d'information **ori-oai-info**

<http://listes.univ-rennes1.fr/wws/info/ori-oai-info>

Ce travail est mis à disposition sous une licence Creative Commons

Vous êtes libres

De reproduire, distribuer et communiquer cette création au public

De modifier cette création



Cette création est mise à disposition selon le Contrat Paternité-NonCommercial-ShareAlike 2.5 disponible en ligne <http://creativecommons.org/licenses/by-nc-sa/2.5/>

- Remarque : Les transparents présentés ici ont été réalisés par :
 - Rosa-Maria Gomez (INSA de Lyon)
 - Yohan Colmant (Université de Valenciennes)
 - Raymond Bourges (Université de Rennes 1)
 - Monique Joly (INSA de Lyon)
 - Alain Mayeur (Université de Valenciennes)
 - François Jannin (EINSEEIHT – INP de Toulouse)
 - Nolwen Huet (INSA de Lyon)