

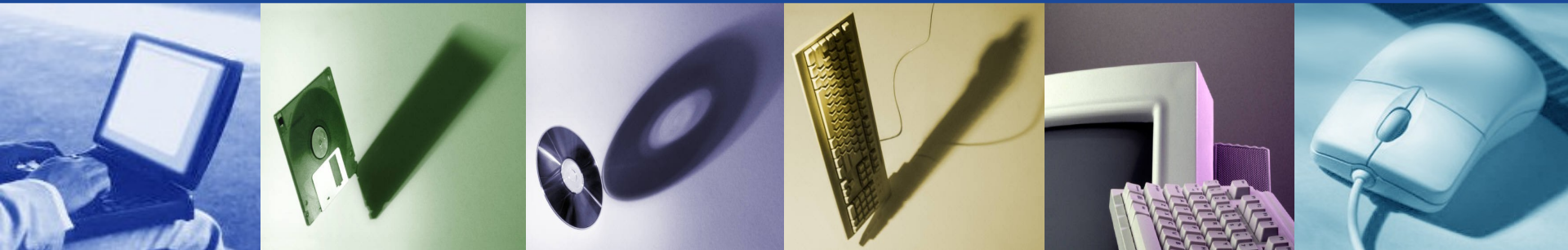


Projet ORI-OAI

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

Réseau de portails **OAI**

Stage Logiciels libres, ENSSIB
Villeurbanne, 23 octobre 2008



Introduction

Contour fonctionnel

Concepts

Mise en oeuvre

Évolutions et projets connexes

Bilan

Introduction

■ Les établissements universitaires

- Croissance de la création numérique institutionnelle (scientifique, pédagogique, administrative, documentaire)
- 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

Établissements



UNT



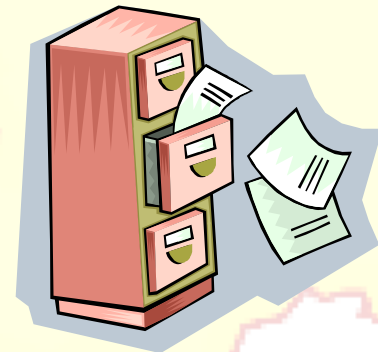
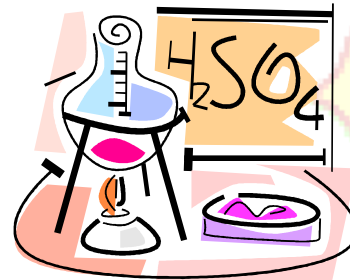
Avec le soutien du MESR



- Équipe de pilotage avec des représentants
 - des établissements
 - des UNT
 - des SCD

■ Équipe fonctionnelle

- 1 coordinatrice
- 10 membres Ressources pédagogiques numériques
- 8 membres Thèses
- 7 membres Archives ouvertes
- 8 membres Documents administratifs
- ...



■ Équipe technique

- 1 coordinateur
- 1 conseiller technique
- 5 développeurs
 - 2 personnes Rennes 1
 - 1 personne INP Toulouse
 - 1 personne SICD Toulouse
 - 2 personnes Valenciennes



- Un projet commun alimenté de l'expérience de chacun
- Une démarche d'ouverture et d'échange
 - Diffusion des sources
 - Site Web
 - Listes de diffusion
- Financement coopératif par les établissements

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** articulées 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**

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

- 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**

Projet ORI-OAI

Insertion de ORI-OAI dans l'établissement

■ Ressources d'enseignement

- ressources pédagogiques multimédia, QCM, exercices, photocopiés numériques, ...

■ Travaux de recherche

- publications, pré-publications, rapports
- Thèses, mémoires de Master

■ Documents administratifs

■ Ressources documentaires et bibliothécaires

- revues, livres électroniques



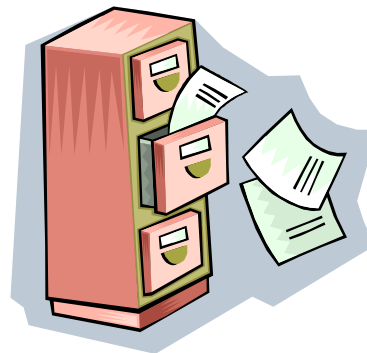
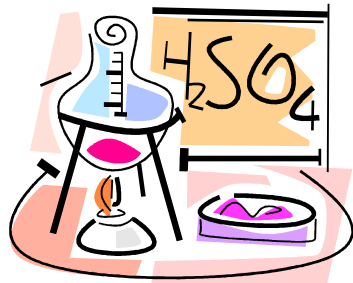
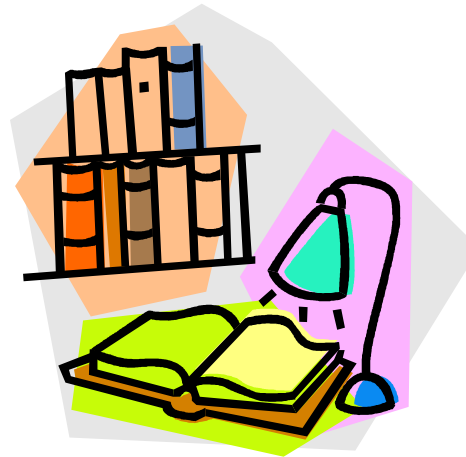
- Patrimoine pérenne pouvant se partager
 - Avec les acteurs de l'établissement
 - Avec les partenaires des réseaux locaux, nationaux et internationaux de l'établissement
 - Avec les internautes
 - En respectant les règles du numérique et les standards
 - Droit d'auteur
 - Interopérabilité des systèmes (OAI/SOAP)
 - Normes d'échange (schémas XML)
 - Vocabulaire commun

- Pourquoi intégrer ORI-OAI dans un établissement?
 - 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



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



- 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:
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 (LMS, SI recherche, HAL, application de la scolarité, ...)
 - Référentiel des documents numériques pour le SI (plateforme pédagogique, moteur documentaire, sites web de communication, ...)

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

- 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

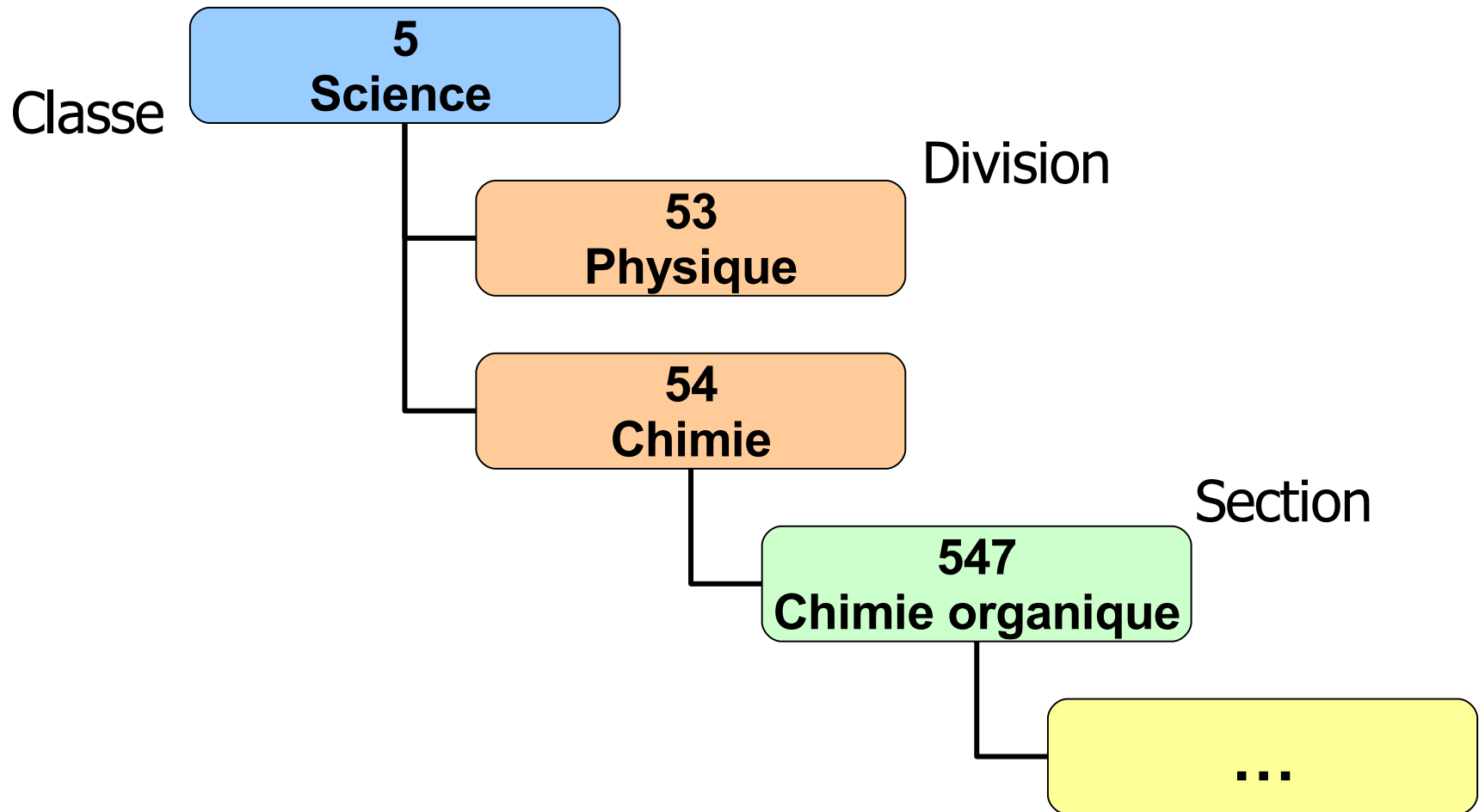
Classification

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

- Objectif ?
 - Échange et partage, interopérabilité
 - Classification internationale
 - Couvre l'ensemble des domaines de la connaissance
 - interopérable avec les applications qui utilisent la Dewey (BU, Thèses)
 - Utilisation arborescente
 - 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** » spécifique pour l'indexation et pour la recherche thématique

Quelle interaction entre la classification locale et la classification Dewey ?



Vocabulaire des Communautés UNIT



■ Classification UNIT

- élaborée par les communautés UNIT
- points d'entrée thématiques sur les ressources numériques

- Objectif ?
 - Permettre l'interaction entre une classification locale et la classification Dewey

- 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

Université
de Valenciennes
et du Hainaut-Cambrésis



Sciences, techniques, sport

Génie mécanique et énergétique

• 621*

...

...

Projet ORI-OAI

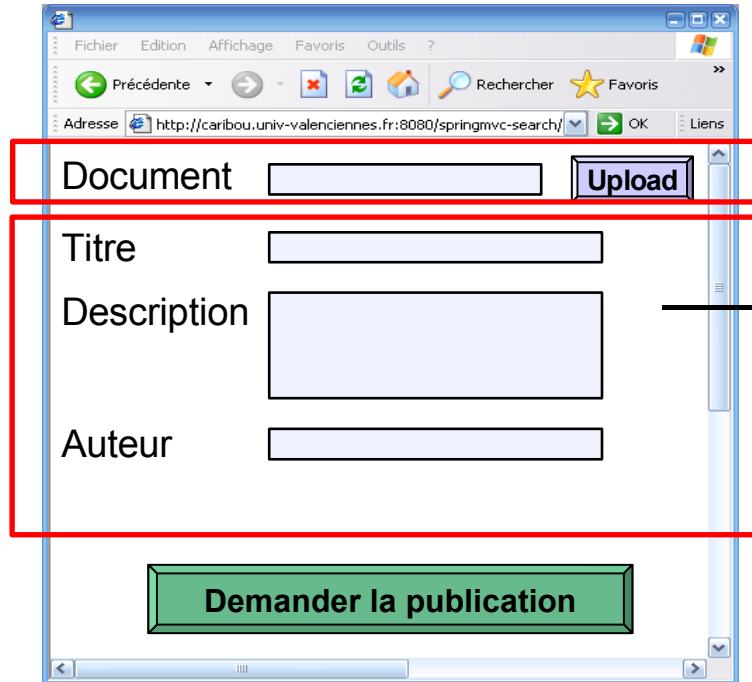
Principe du référencement et de la diffusion

Procédure de publication (*étape auteur*)

Doc



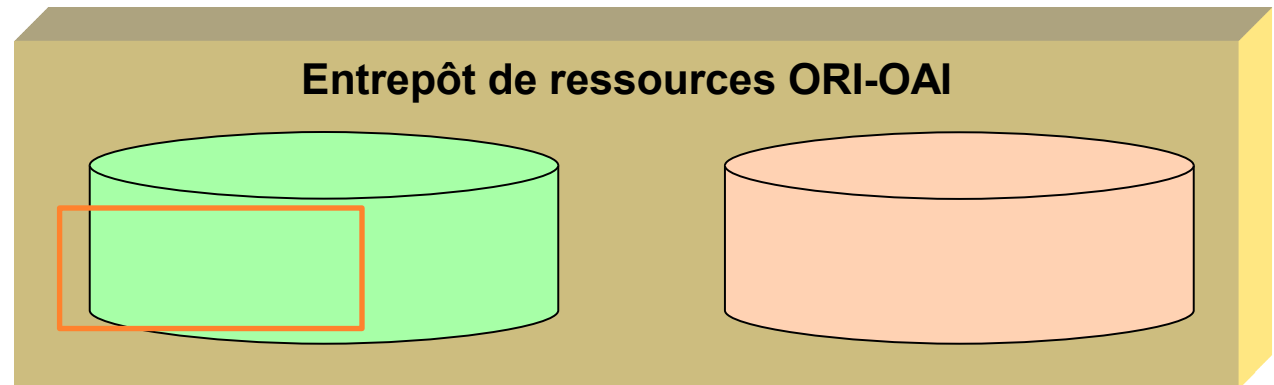
Auteur



Étape de saisie
auteur

Metadonnées

À publier

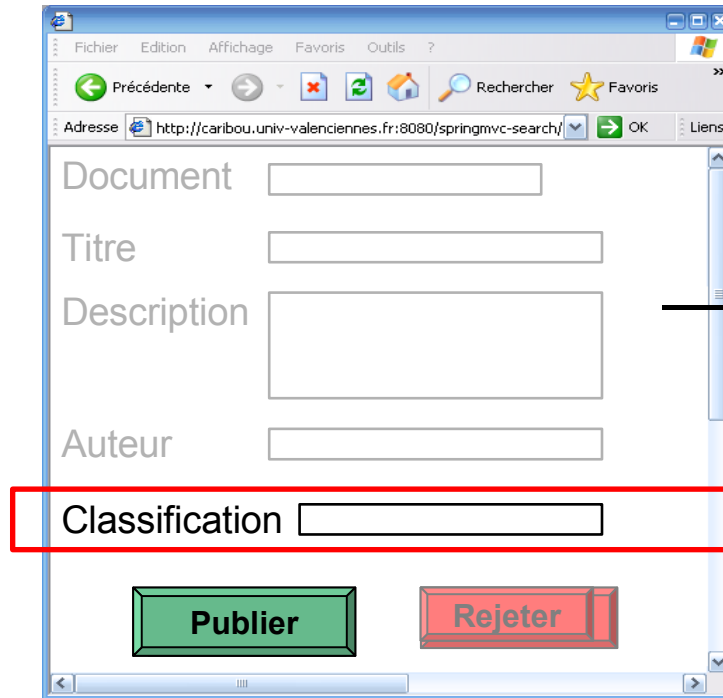


Procédure de publication *(étape bibliothécaire)*



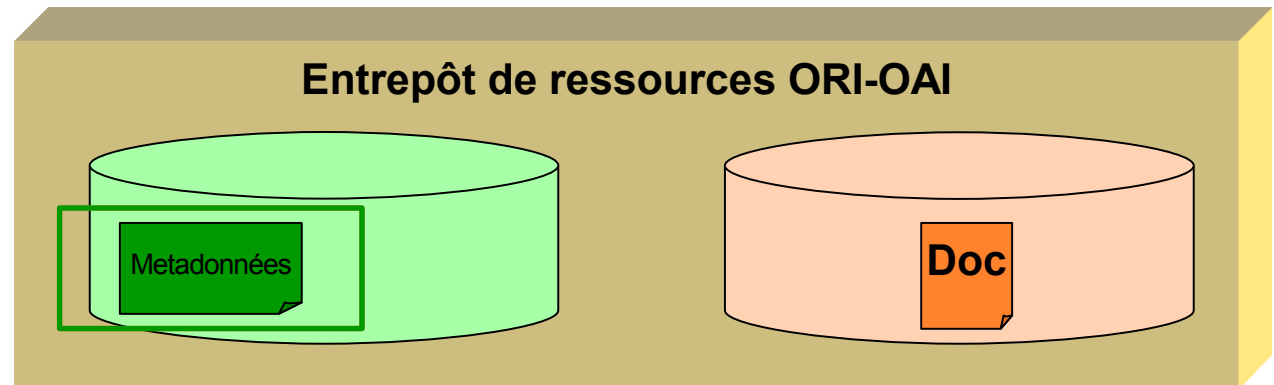
Bibliothécaire

Étape de saisie
bibliothécaire

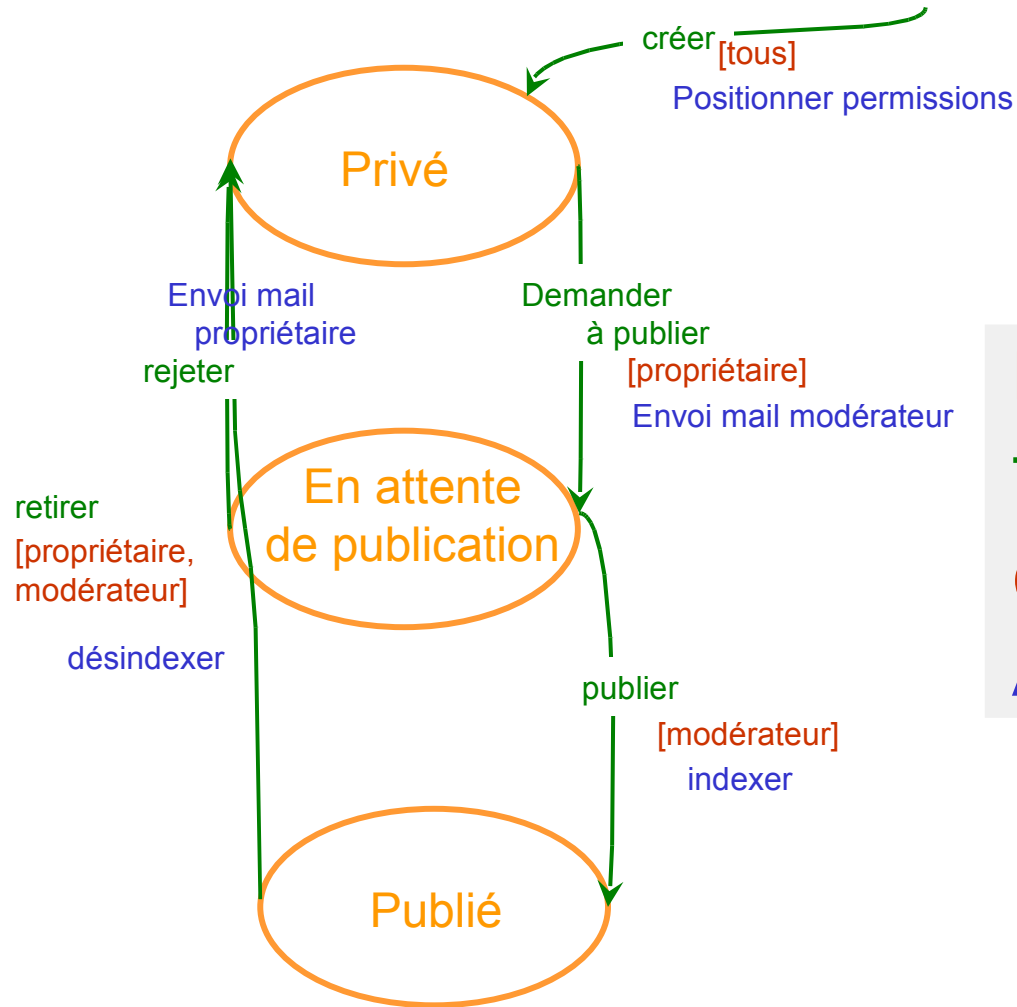


Metadonnées

Publié



Notion de workflow: chaîne de publication



Schématisation du processus :
diagramme Etats / Transitions

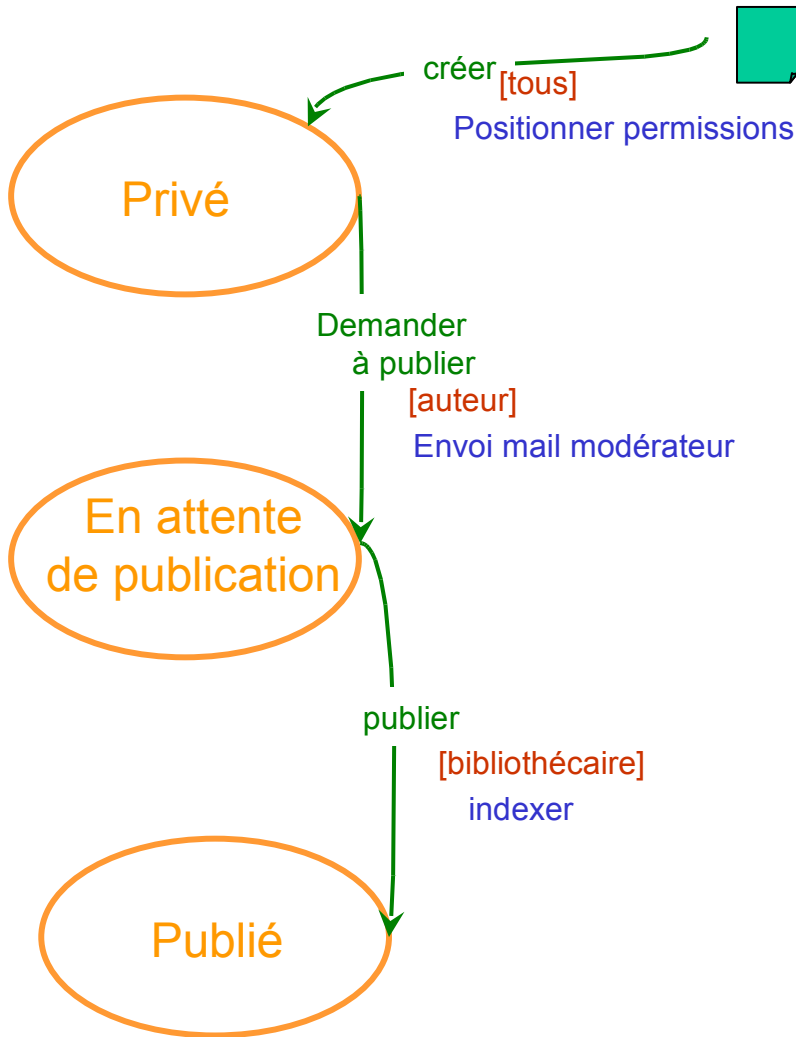
Etats (empruntables par la fiche)

Transitions (changements d'états)

Conditions (pour les transitions)

Actions (lors de transitions)

Notion de workflow: chaîne de publication



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

Etats

Transitions

Conditions

Actions

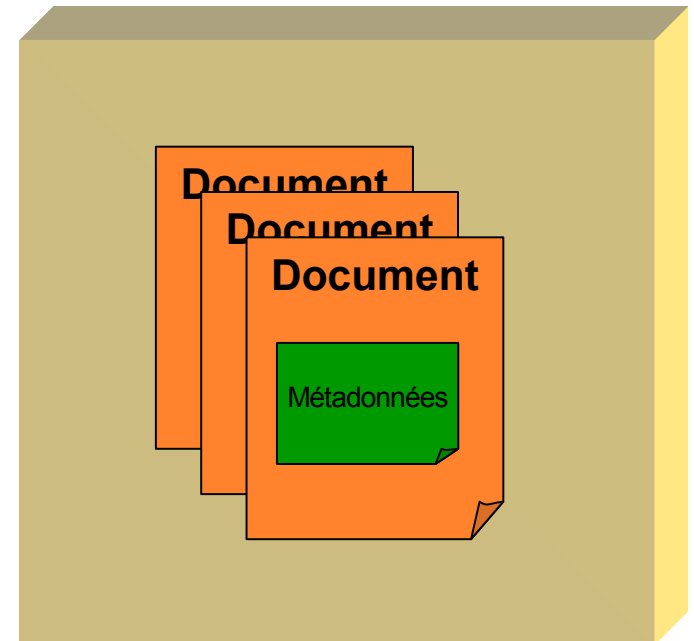
- Centralisation et automatisatisation des vocabulaires
 - Répartition en réseau des vocabulaires avec tolérance à la panne (cache intelligent)
 - Utilisés en amont (workflow) et en aval (recherche, OAI)
 - Chaque vocabulaire est partageable par x établissements, UNT/UNR
 - Vocabulaires de référence fermés et **statiques** avec peu de fluctuations (taxonomies, liste de mots-clés fermée)
 - Vocabulaires **dynamiques** générés automatiquement
 - depuis un annuaire LDAP ou Active Directory
 - d'après les valeurs déjà indexées (auteurs, mots-clés libres)
 - d'après des référentiels existants (bases de données...)

Projet ORI-OAI

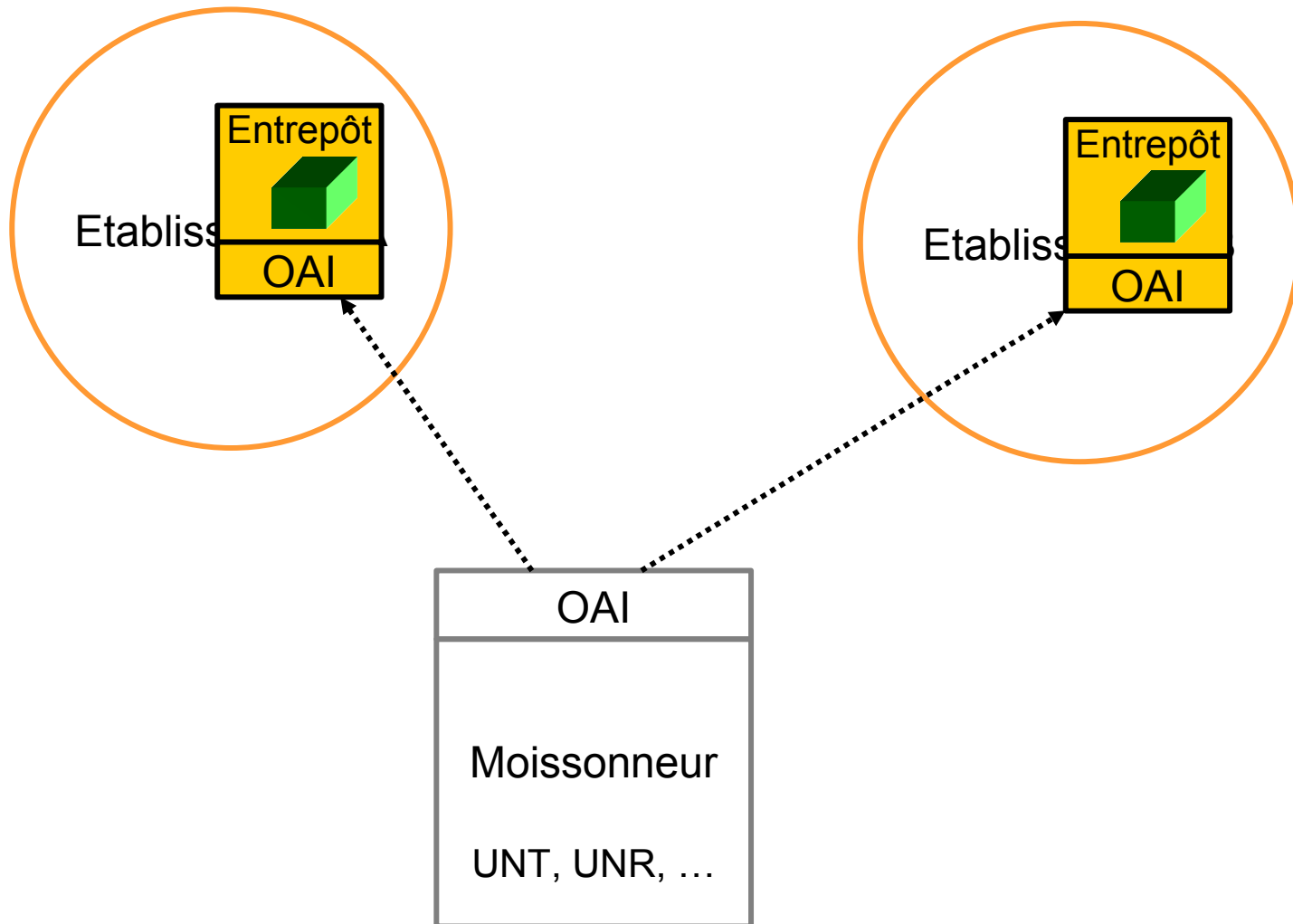
Principe des échanges

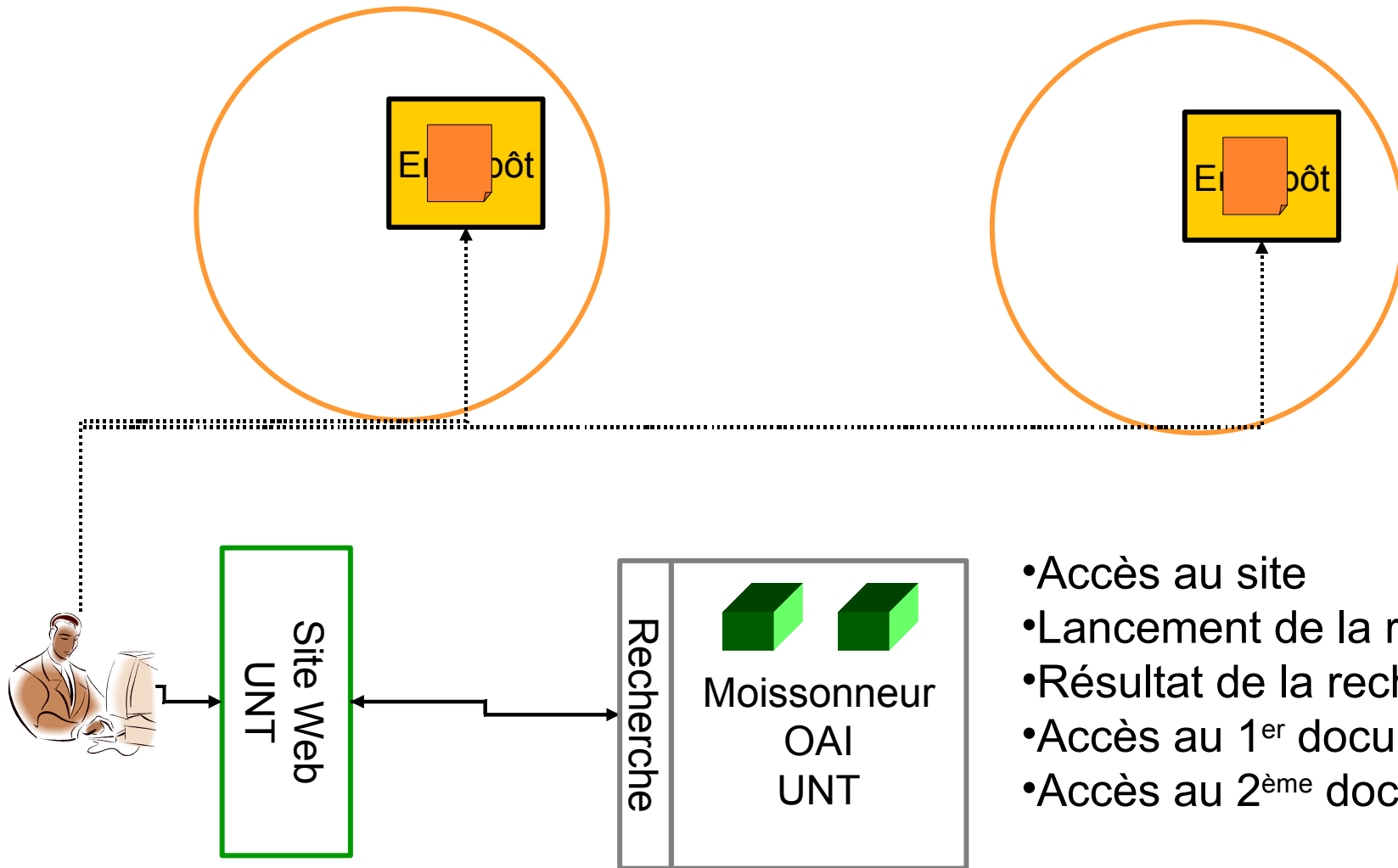
■ Qu'est-ce qu'un entrepôt ORI-OAI ?

- Stockage des ressources numériques
- Métadonnées sur les documents
 - LOM
 - Dublin Core
 - Etc.



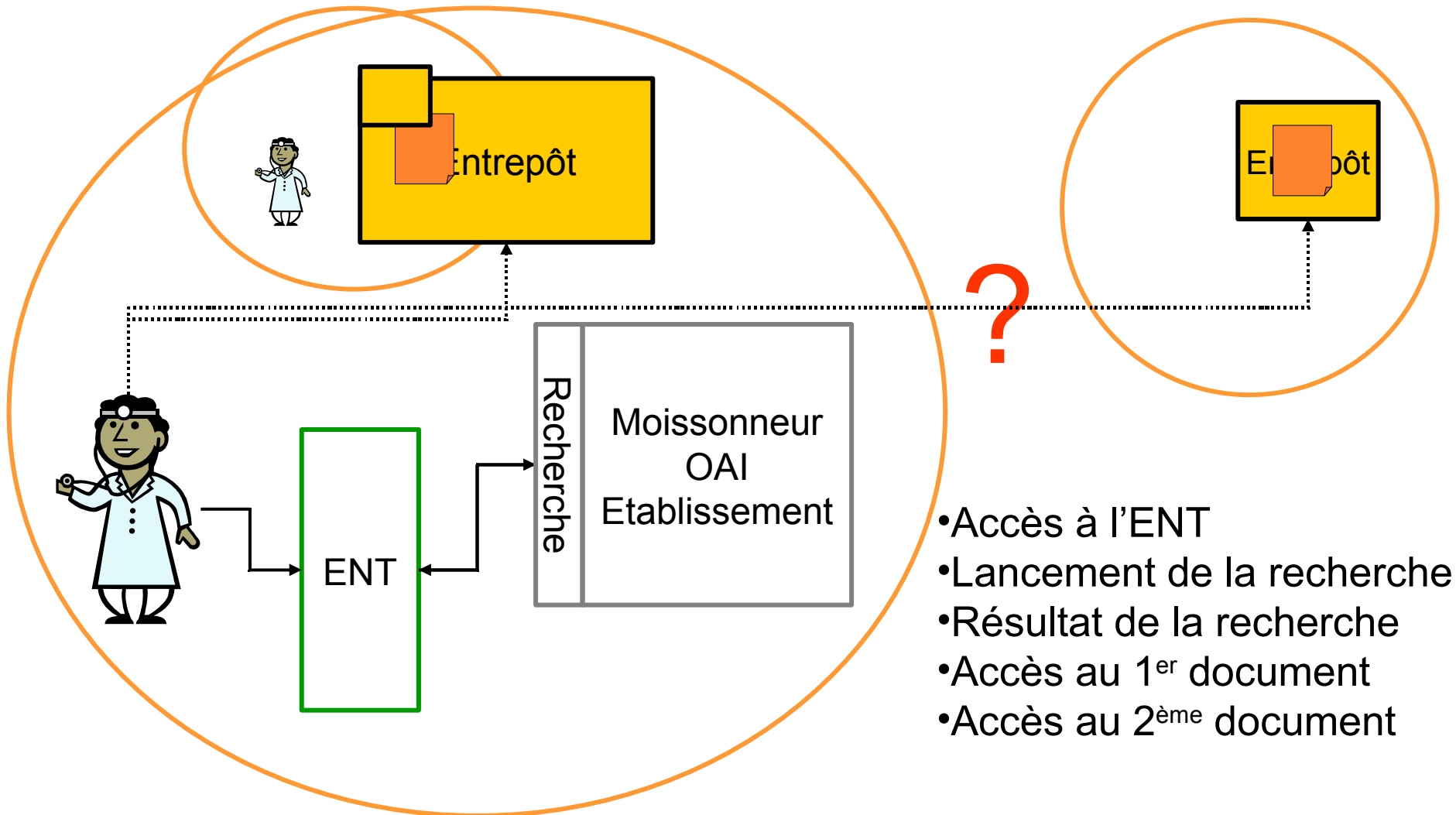
- Protocole informatique d'échange de données
- Permet la collecte et le partage de références
- Fait intervenir deux types d'acteurs:
 - Fournisseur de données (data provider) :
 - *met en forme les métadonnées (au moins en DC)*
 - *les stocke dans un entrepôt (repository)*
 - *définit éventuellement 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*

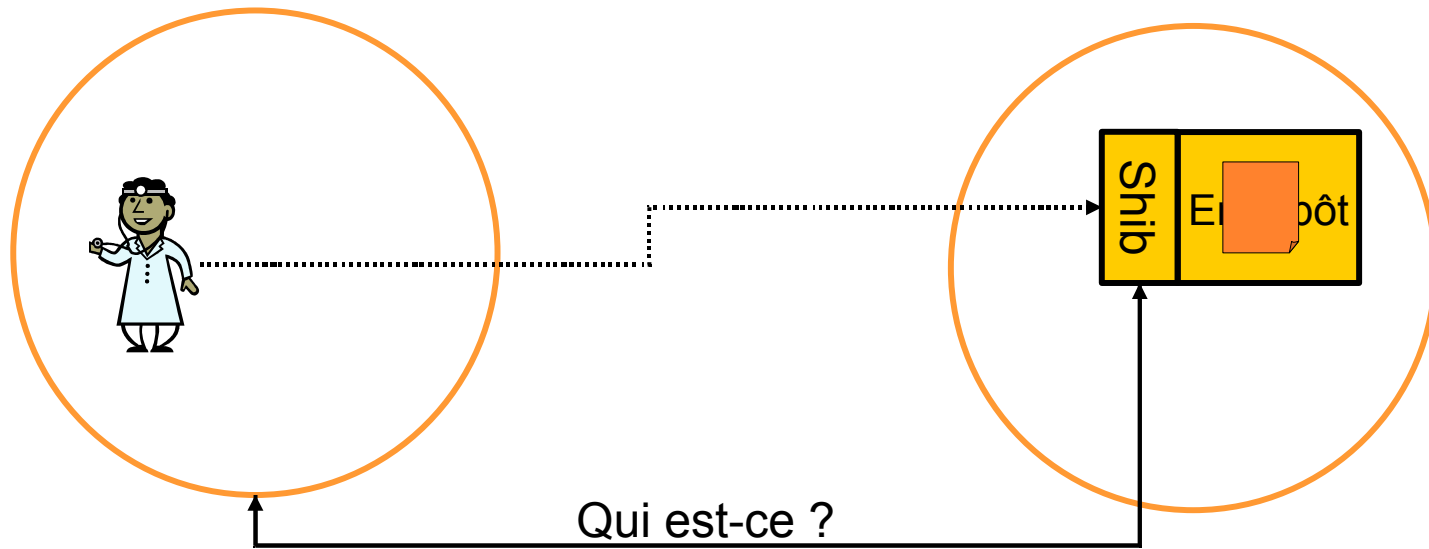




- Accès au site
- Lancement de la recherche
- Résultat de la recherche
- Accès au 1^{er} document
- Accès au 2^{ème} document

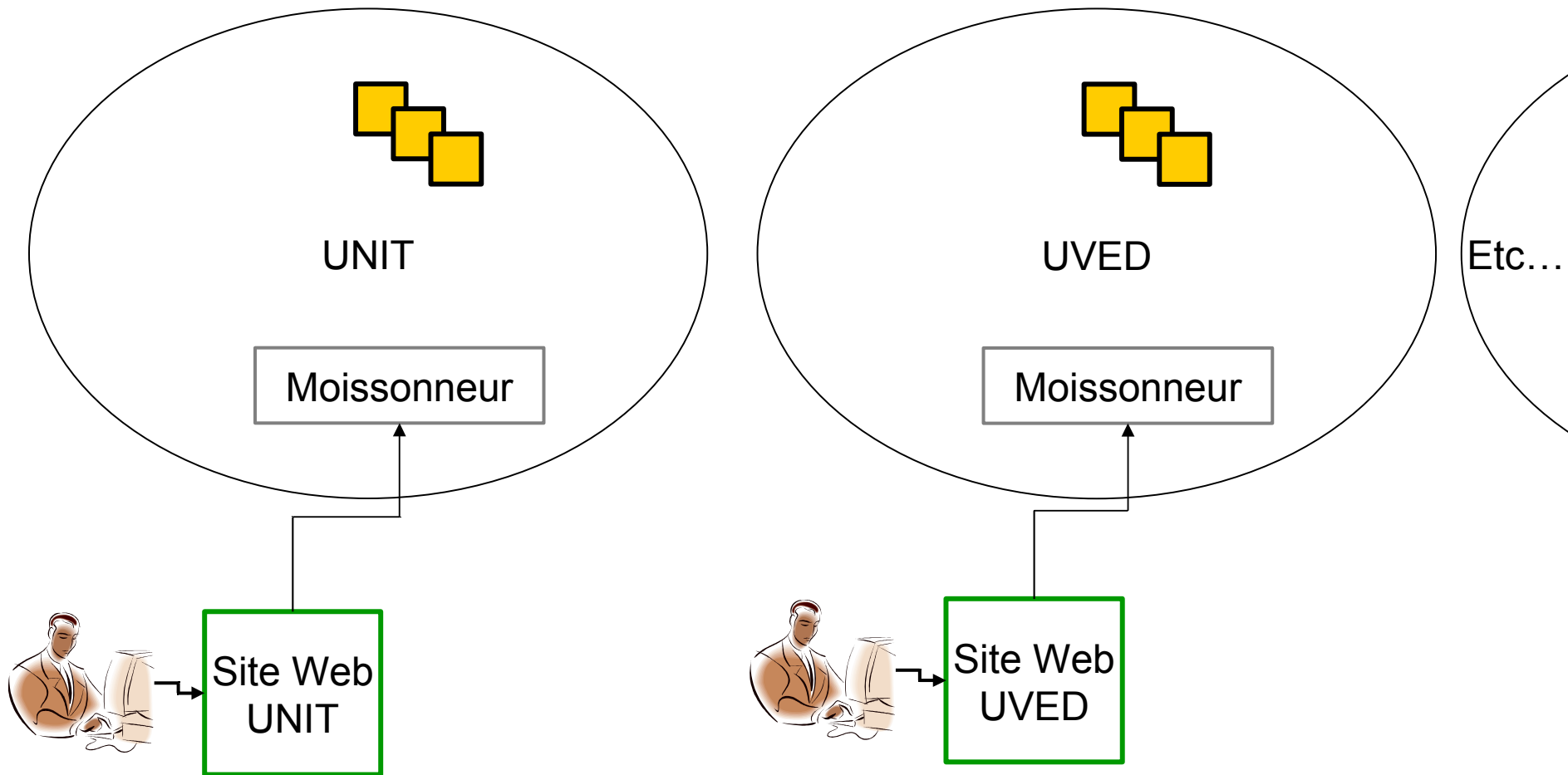
OAI pour l'utilisateur d'un établissement

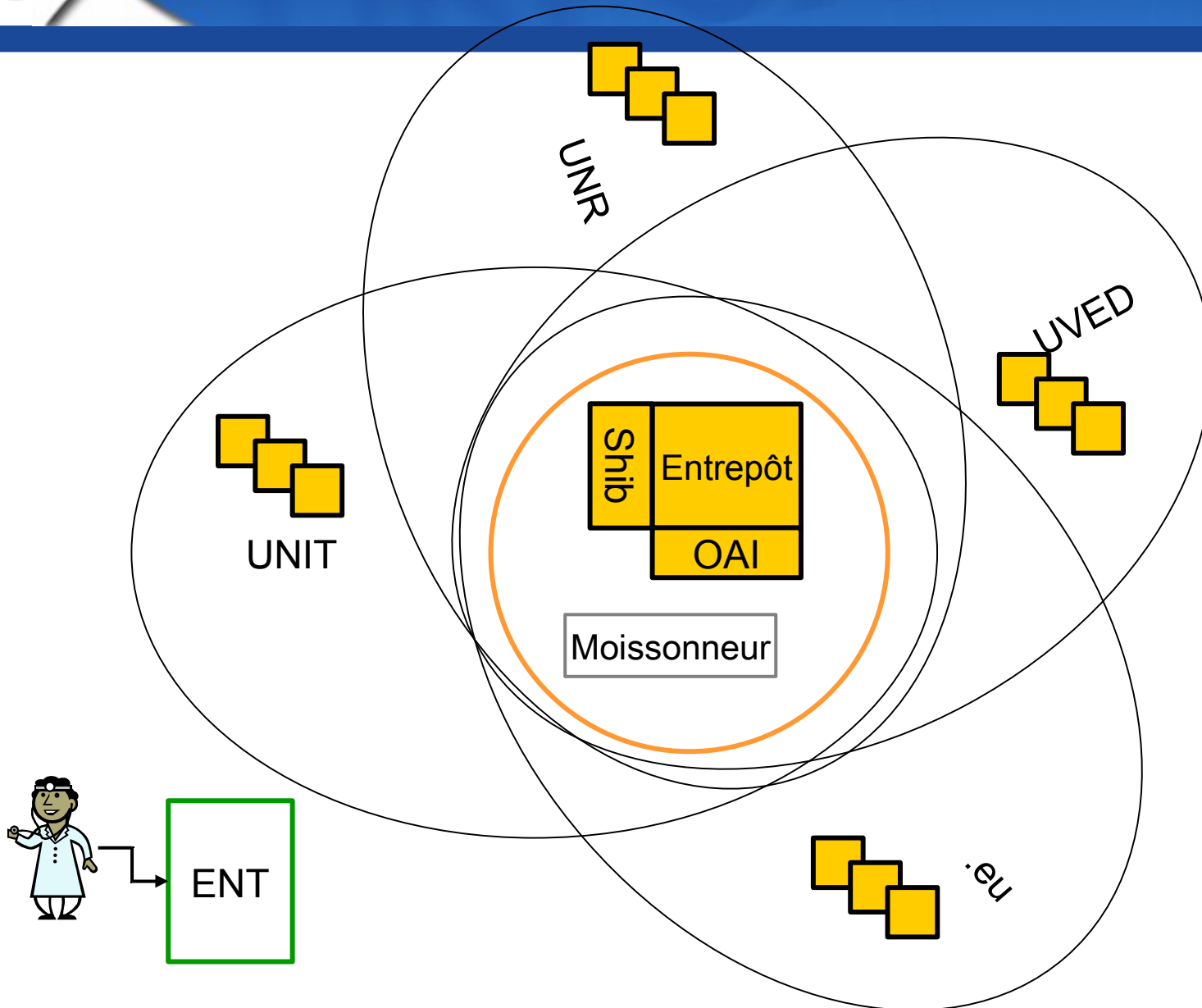




Un étudiant en médecine 2^{ème} année

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





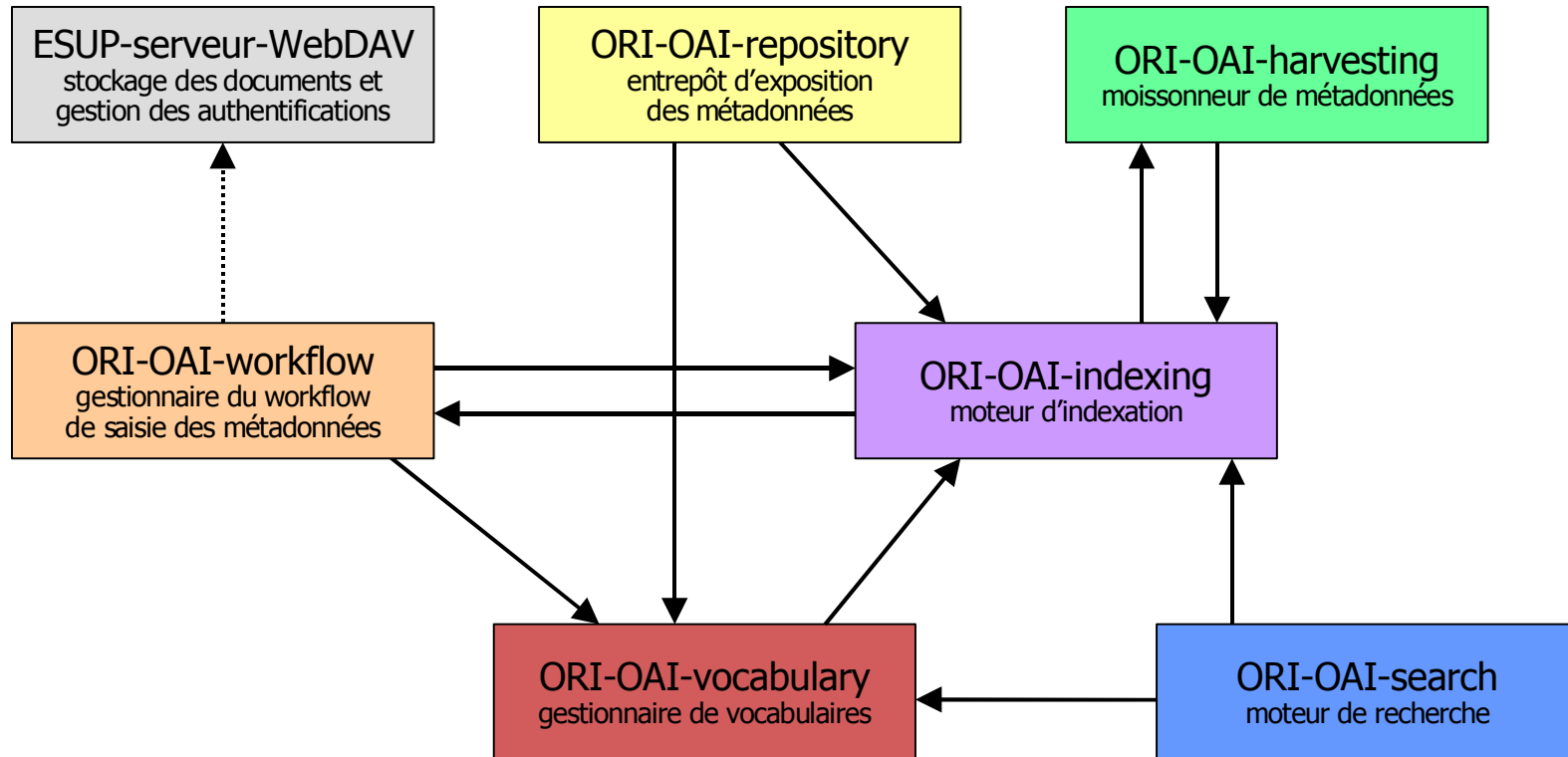
Mise en oeuvre

Projet ORI-OAI

Architecture

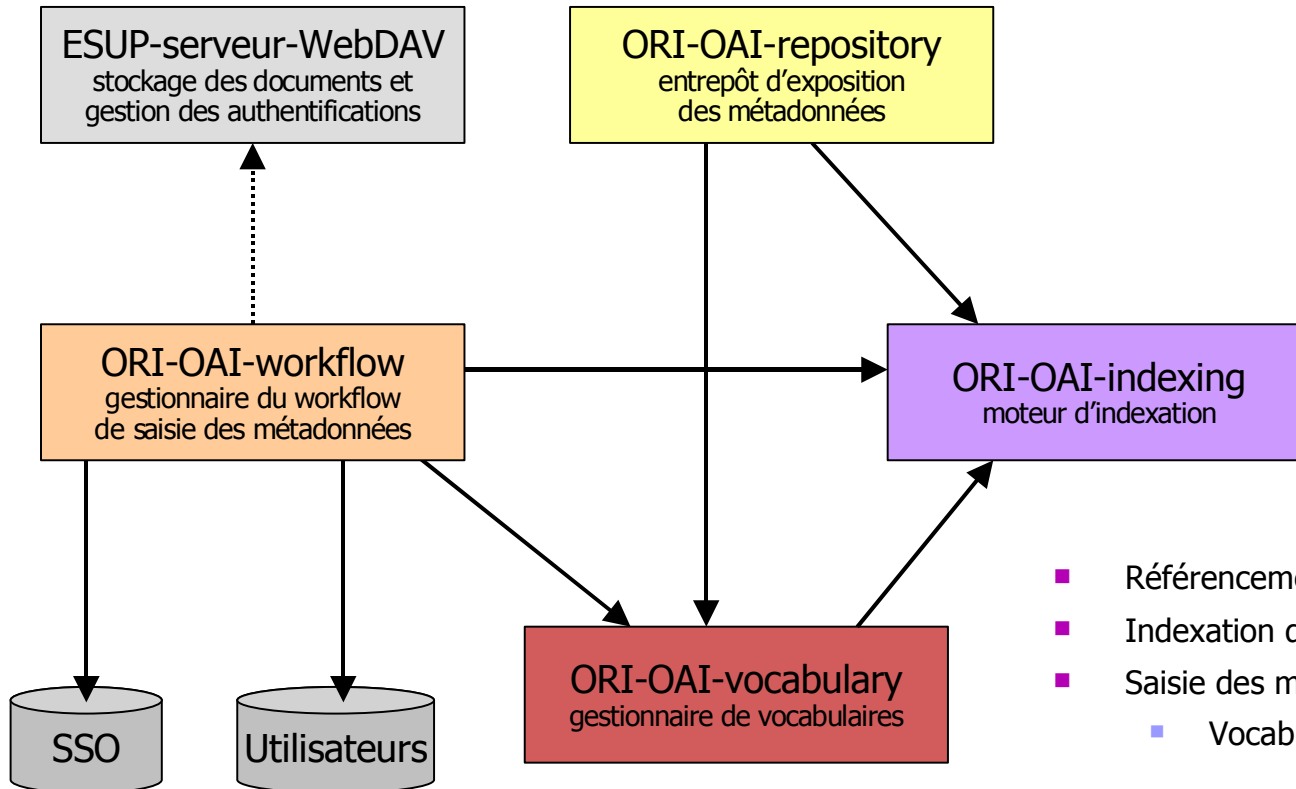
- 7 modules
 - Indépendants
 - Ayant chacun un rôle précis
 - Tous interopérables
 - Dialoguent entre eux par Web Services

- 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



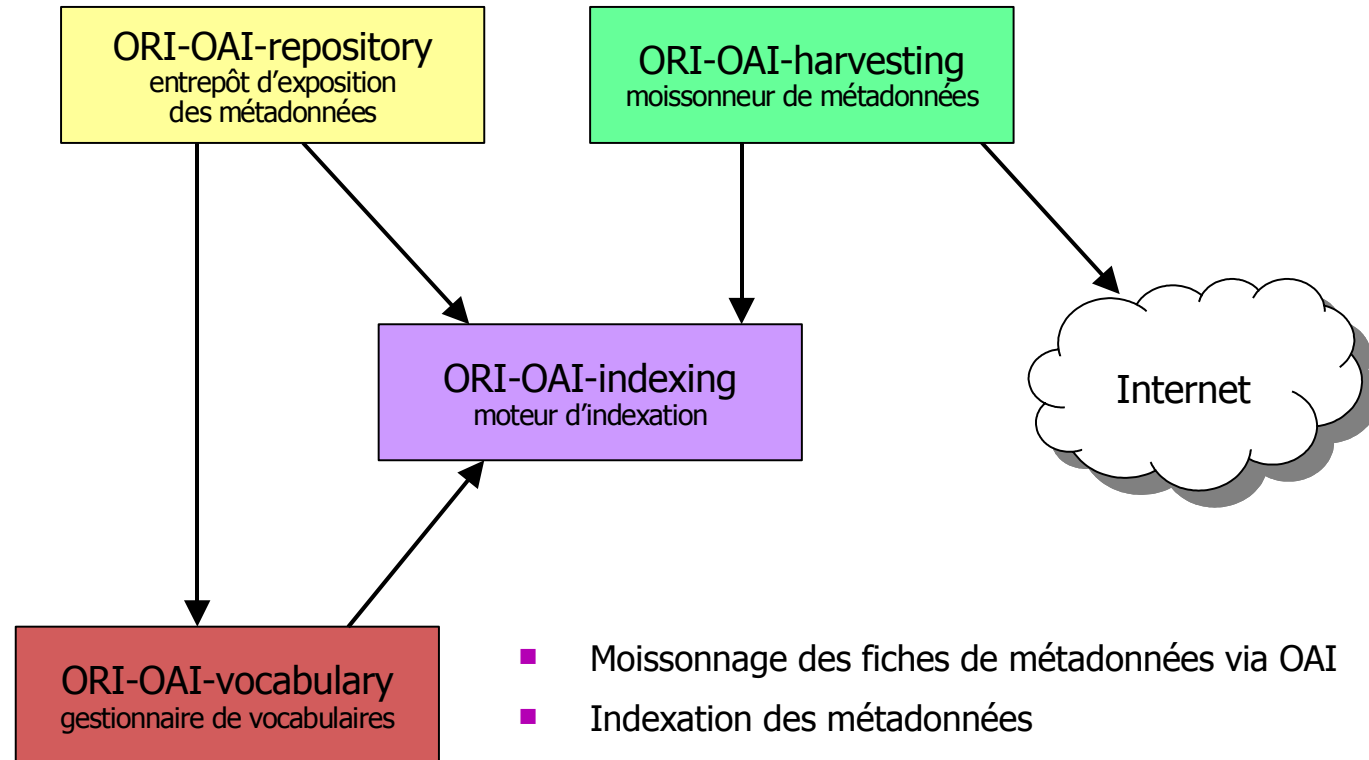
- 7 modules
- interopérables

Cas 1: producteur de données



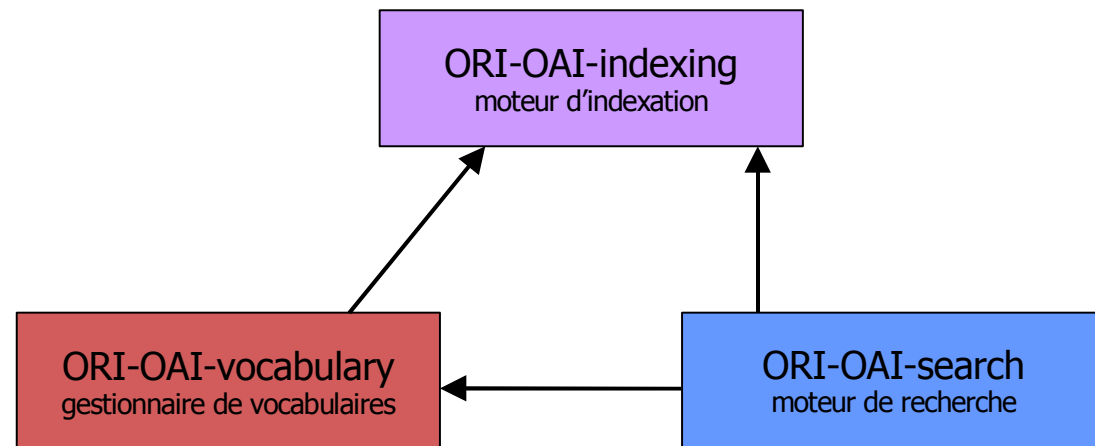
- Référencement du lien vers le(s) document(s)
- Indexation des métadonnées
- Saisie des métadonnées via vocabulaires
 - Vocabulaires dynamiques basés sur l'index
- Exposition OAI (optionnelle)
 - Exposition des fiches par requêtes sur l'index
 - Génération de sets OAI via le vocabulaire

Cas 2: Agrégateur de données



- 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

- 1.0 disponible depuis le 15 octobre 2007
- 1.1 depuis juin 2008
- 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 est mis en place (ou en cours de test) par :
 - UNIT
 - UNISCIEL
 - UNR Nord-pas-de-Calais
 - Université de Rennes 1
 - Université de Valenciennes
 - INSA de Lyon
 - INP de Toulouse / ENSEEIHT
 - Université des Antilles
 - Université d'Artois
 - Université de Bordeaux 1
 - Université de Brest
 - Université de Bretagne Sud
 - Université de Limoges
 - Université du Mans
 - Université de Nancy 2
 - Université de Paris 5
 - Université de Paris 10
 - Université de Paris Sud
 - Université de Rennes 2
 - Université de Versailles
 - IUFM de Bretagne
 - Ecoles de Saint-Cyr Coëtquidan
 - ENS Lyon
 - Campus Virtuel Marocain
 - Université virtuelle de Tunis
 - etc.

Évolutions et projets connexes

■ Version 1.5.0

■ 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
 - Statistiques
 - De recherche
 - D'accès aux documents

■ Prochaines versions

- Référencement et publication dans l'archive de l'établissement
 - des thèses avec le format TEF
 - des publications de la recherche
- 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)