

Utilisation du protocole OAI-PMH pour diffuser ses notices

Lucie Gay, ENS Paris-Saclay | lucie.gay@ens-paris-saclay.fr

école —————
normale —————
supérieure —————
paris — saclay —————

Présentation du protocole OAI-PMH

Le protocole OAI-PMH

- « Open Archives Initiative - Protocol for Metadata Harvesting »
 - Protocole international, libre d'utilisation
 - Version stable depuis 2002
-
- Facilite l'interopérabilité des ressources, sans duplication
 - Créé pour la diffusion des métadonnées d'articles scientifiques en accès ouvert

Le protocole OAI-PMH

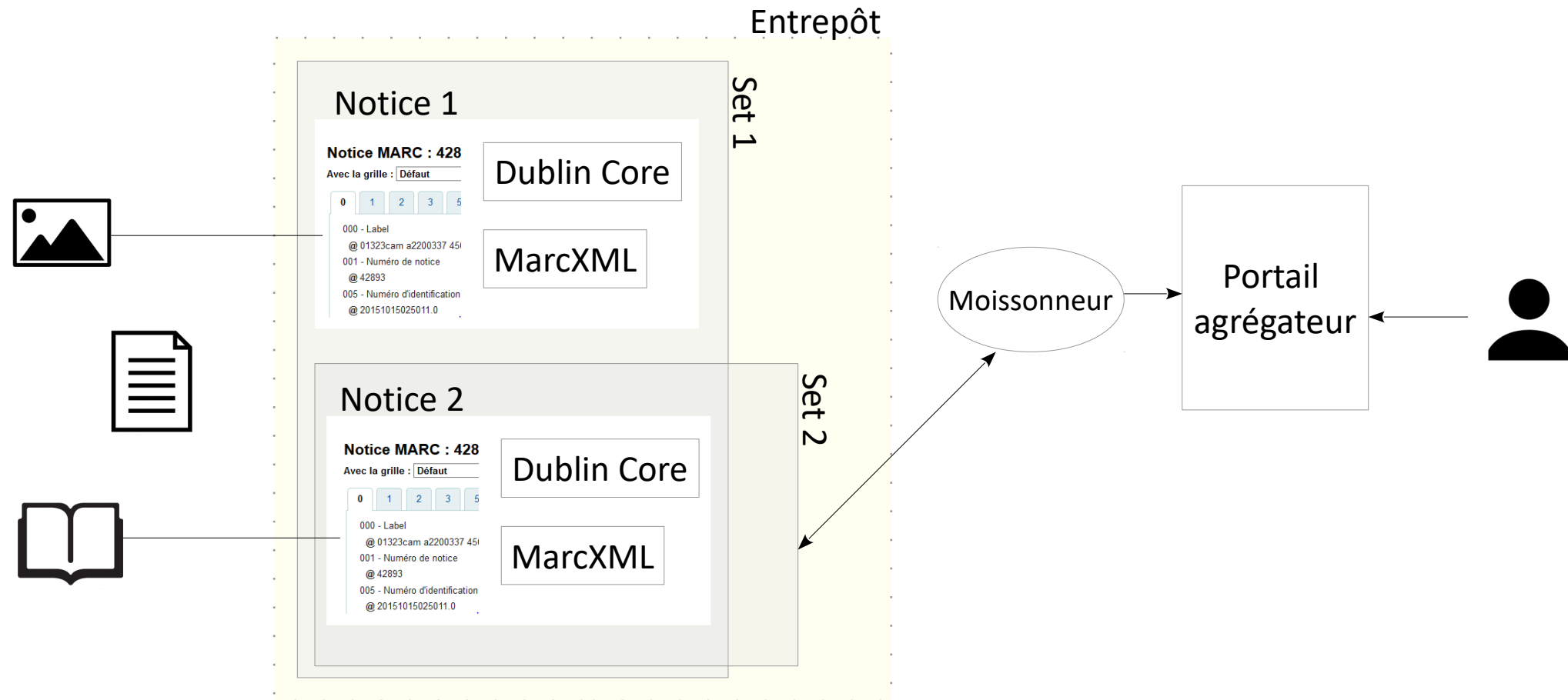
- Moyen d'échanger des métadonnées entre plusieurs institutions, afin de multiplier les accès
- Utilisation d'un format de description générique :
Dublin Core
- Met en jeu deux acteurs :
 - le **fournisseur de données** : celui qui expose ses données
 - le **fournisseur de services** : celui qui récupère (moissonne) les données

Fournisseur de données

- Donne accès son catalogue,
à travers un entrepôt OAI
 - Sets : sous-ensembles du catalogue correspondant à certains critères
 - Métadonnées au format Dublin Core
 - Autres formats possibles, comme le MarcXML

Fournisseur de services

- Appelé aussi moissonneur
- Collecte les métadonnées mises à disposition par les fournisseurs de données
- Intègre les métadonnées dans ses propres outils
- Propose une interrogation de sa base
 - Les entrepôts d'origine ne sont pas sollicités
 - Lien vers document primaire, accessible sur le serveur du fournisseur de données



Ressources
physiques
ou numériques
→ objet décrit

Notices décrivant la ressource et
leurs enregistrements au format XML
→ métadonnées à moissonner

Points forts

- Échange les métadonnées, pas les objets
- Format de description générique (Dublin Core), possibilité d'utiliser des formats spécifiques
- Pour le fournisseur de données
 - Meilleure visibilité de ses données
- Pour le fournisseur de services :
 - Agrégation de métadonnées à partir de sources différentes
 - Gestion simple et automatisée des mises à jour, choix de la fréquence

Utiliser un entrepôt OAI pour exposer son catalogue Koha

Mise en place de l'entrepôt OAI

→ Préférences système > Web service

▼ OAI-PMH

Préférences ▲	Valeur
OAI-PMH	<input type="text" value="Activer"/> le serveur OAI de Koha OAI-PMH .
OAI-PMH:archiveID	Identifier les enregistrement de ce site avec le préfixe : <input type="text" value="KOHA-OAI-TEST"/> : .
OAI-PMH:AutoUpdateSets	<input type="text" value="Activer"/> la mise à jour automatique des sets OAI-PMH quand une notice bibliographique est ajoutée ou modifiée
OAI-PMH:ConfFile	Fichier YAML de configuration du serveur OAI de Koha : <input type="text" value="/home/koha/etc/oai/oai.yaml"/> Si vide, le serveur OAI fonctionne en mode normal, autrement il opère en mode étendu. En mode étendu, il est possible de définir d'autres formats que le marcxml ou le Dublin Core. Le fichier YAML liste les formats de métadonnées disponibles et les fichiers XSL utilisés pour générer ces formats à partir des notices marcxml.
OAI-PMH:DeletedRecord	web_services.pref##OAI-PMH:DeletedRecord# table "deleted biblio" de Koha <input type="text" value="web_services.pref##OAI-PMH:DeletedRecord# Ne sera jamais vidé ou tronqué (persistant)"/> web_services.pref##OAI-PMH:DeletedRecord#.
OAI-PMH:MaxCount	Renvoyer les résultats par lot de <input type="text" value="50"/> notices à la fois en réponse aux requêtes ListRecords et ListIdentifiers.

Mise en place de l'entrepôt OAI

- Activer le serveur OAI
- Donner un préfixe pour l'identifiant de la notice dans l'entrepôt
 - Préfixe:biblionumber (ex : KOHA-OAI-TEST:60868)
- AutoUpdateSets
- ConfFile : pour définir les formats des métadonnées
 - si vide → uniquement Dublin Core
 - le fichier *oai.yaml* doit contenir la liste des formats dans lesquels seront proposées les métadonnées

Mise en place de l'entrepôt OAI

- Fichier *oai.yaml*

```
format:
  marcxml:
    metadataPrefix: marcxml
    metadataNamespace: http://www.loc.gov/MARC21/slim http://www.loc
    schema: http://www.loc.gov/MARC21/slim http://www.loc.gov/standa
    include_items: 1
  oai_dc:
    metadataNamespace: http://www.openarchives.org/OAI/2.0/oai_dc/
    metadataPrefix: oai_dc
    schema: http://www.openarchives.org/OAI/2.0/oai_dc.xsd
    xsl_file: /home/koha/src/koha-tmpl/intranet-tmpl/prog/en/xslt/UN
```

Mise en place de l'entrepôt OAI

→ Administration > Configuration des sets OAI

- Nouveau set OAI :
 - SetSpec : code du set
 - SetName : nom lisible du set
- Définir les critères de sélection (*facultatif*) :
 - Si aucun critère → toutes les notices seront dans le set
 - Critères sur champ et sous-champ

Mise en place de l'entrepôt OAI

→ Définir les critères de sélection du set

Champ	Sous-champ	Opérateur	Valeur		
<input type="text"/>	<input type="text"/>	est égal à est égal à ▾	<input type="text"/>	OU	 Supprimer

Mise en place de l'entrepôt OAI

Exemple :

- Set " test_ens "
- Toutes les notices sauf celles avec un code collection = WEB (ressources en ligne)

Champ	Sous-champ	Opérateur	Valeur		
099	t	n'est pas égal à 	WEB	OU	 Supprimer

Mise en place de l'entrepôt OAI

→ Activer l'entrepôt OAI

- Accès au serveur
- Lancer le script *build_oai_sets.pl*
 - src/misc/migration_tools/

```
Summary:  
Set 'test_ens': 60141 biblios  
Updating database... done.
```


Interroger l'entrepôt OAI

- Le moissonneur interroge l'entrepôt OAI avec une requête sous la forme d'une URL construite en deux parties
 - URL de base de l'entrepôt
 - Requête d'interrogation
- La 2^e partie repose sur différents verbes prévus par le protocole

Interroger l'entrepôt OAI

- 6 verbes. Certains fonctionnent avec des arguments (ou attributs).
 - **GetRecord** : afficher un enregistrement en fonction d'un identifiant et d'un format (attributs obligatoires)
 - **Identify** : informations générales sur le serveur
 - **ListIdentifiers** : liste des identifiants et des dates de dernière modification des items disponibles dans la base, dans un format donné (attribut obligatoire)
 - **ListMetadataFormats** : la liste des formats de métadonnées que l'entrepôt est capable de fournir
 - **ListRecords** : liste de tous les items, dans un format donné (attribut obligatoire)
 - **ListSets** : liste des ensembles formés par le fournisseur de données

Interroger l'entrepôt OAI

	From <i>date de début</i>	Until <i>date de fin</i>	MetadataPrefix	Identifiant	Set	Token
Identify						
ListMetadataFormat				facultatif		
ListSets						
ListIdentifiers	facultatif	facultatif	obligatoire		facultatif	exclusif
ListRecords	facultatif	facultatif	obligatoire		facultatif	exclusif
GetRecords			obligatoire	obligatoire		
<i>exemple</i>	<i>2018-01-01</i>	<i>2018-05-31</i>	<i>oai_dc</i>	<i>KOHA-OAI-TEST:60868</i>	<i>test_ens</i>	

Source : Daniel Charnay, *OAI-PMH Le protocole des archives ouvertes*, 2005

Interroger l'entrepôt OAI

- URL catalogue
 - <https://cataloguetest.ens-cachan.fr>
- URL entrepôt OAI
 - <https://cataloguetest.ens-cachan.fr/cgi-bin/koha/oai.pl>

Interroger l'entrepôt OAI

- Exemples de requêtes d'interrogation :
 - `https://cataloguetest.ens-cachan.fr/cgi-bin/koha/oai.pl?verb=Identify`
 - `.../cgi-bin/koha/oai.pl?verb=ListSets`
 - `.../cgi-bin/koha/oai.pl?verb=ListRecords&metadataPrefix=oai_dc`
 - `...?verb=ListRecords&metadataPrefix=marcxml&set=test_ens`
 - `...?verb=GetRecord&identifier=KOHA-OAI-TEST:60876&metadataPrefix=marcxml`