

# Mettre en place un entrepôt OAI-PMH compatible avec le portail FranceArchives

## Table des matières

1.	Qu'est-ce que l'OAI-PMH ?.....	2
2.	Comment fonctionne l'OAI-PMH ?.....	2
3.	L'OAI-PMH sur le portail FranceArchives .....	2
3.1	Fournisseur de service.....	2
3.1.1	Adapter l'encodage à un environnement distant .....	3
3.1.2	Transmission d'identifiants fiables.....	4
3.1.3	Sécuriser l'accès à l'entrepôt.....	5
3.1.4	Paramétrer l'entrepôt .....	5
3.1.5	Moissonnage différentiel .....	5
3.1.6	Automatisation des suppressions .....	5
3.2	Fournisseur de données .....	5
3.2.1	Requêtes et formats supportés.....	5
3.2.2	Exemples de requêtes .....	6

## 1. Qu'est-ce que l'OAI-PMH ?

Le protocole OAI-PMH permet d'échanger des métadonnées sur internet, de façon automatisée. Il permet d'une part à un fournisseur d'exposer les métadonnées de ses ressources dans des entrepôts, d'autre part de collecter les métadonnées présentes dans ces entrepôts, afin de les exposer sur une ou plusieurs applications.

L'OAI-PMH constitue une extension au protocole standard du web : HTTP, en définissant six verbes d'interrogations spécifiques.

## 2. Comment fonctionne l'OAI-PMH ?

Le protocole implique deux acteurs :

- **Le fournisseur de données** : il expose les métadonnées contenues dans un entrepôt.
- **Le fournisseur de services** : il moissonne un ou plusieurs entrepôts en utilisant les interfaces exposées par le fournisseur de données, afin d'offrir aux utilisateurs des interfaces de recherche ou de navigation.

Dans un entrepôt OAI-PMH, on appelle « **record** » chaque ressource stockée correspondant à une notice. Une notice peut être disponible selon plusieurs formats de données, *a minima* il est recommandé de permettre son exposition en **Dublin Core simple** (voir la [recommandation de description des archives en DC du portail FranceArchives](#)). Il est par ailleurs vivement conseillé d'utiliser des vocabulaires plus riches et plus spécifiques en fonction des besoins. Dans notre cas, il convient d'envisager la description des métadonnées archivistiques dans un entrepôt en utilisant l'**EAD**. En EAD, un *record* correspond à un instrument de recherche encodé dans son intégralité, quel que soit sa longueur.

Les notices peuvent être regroupées au sein d'ensembles, appelés « **set** ». Cela permet notamment la moisson d'un ensemble spécifique par le fournisseur de services.

Pour en savoir plus :

- Présentation de l'OAI-PMH sur Open Archives  
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

## 3. L'OAI-PMH sur le portail FranceArchives

### 3.1 Fournisseur de service

Le portail moissonne actuellement de nombreux entrepôts de services partenaires. Voici la liste des formats supportés par le moissonneur :

- XML-EAD (metadataPrefix=oai\_ead ou ead)
- XML Dublin Core (metadataPrefix=oai\_dc)
- XML Nomina (metadataPrefix=nomina)

Le moissonnage cible un service particulier, il est réalisé avec le verbe « ListRecords », un paramètre « Set » peut être adjoint à la requête mais il n'est pas obligatoire.

Le moissonneur a été développé à partir de l'outil open source Sickle (<https://pypi.org/project/Sickle/>).

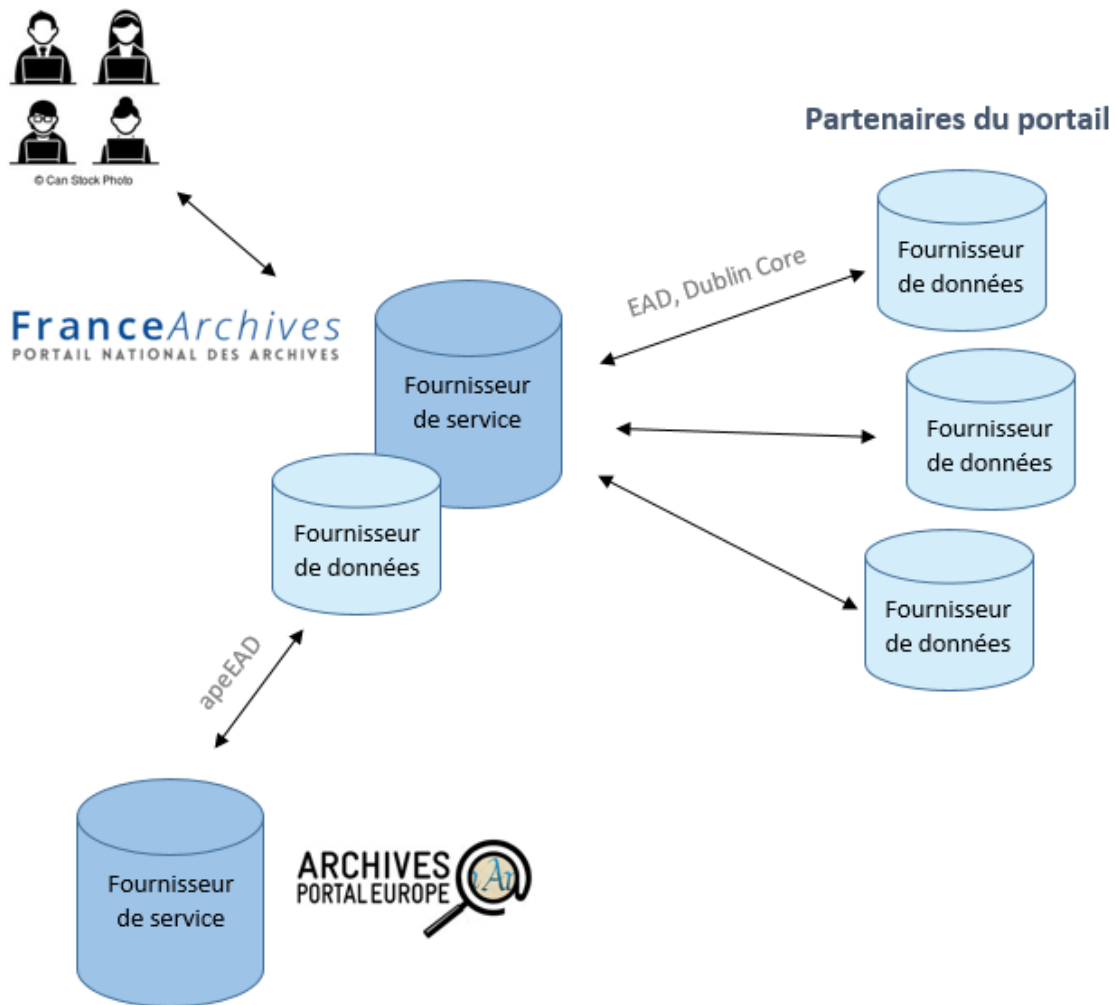


Figure 1 Chaîne de transmission des données sur FranceArchives

### 3.1.1 Adapter l'encodage à un environnement distant

Le portail FranceArchives est un agrégateur de données, destiné à renvoyer vers les sites sources pour la consultation des instruments de recherche et des images numérisées. La capacité du portail à assurer le rebond vers les sites sources dépend de la nature des liens fournis au sein des fichiers transmis par les contributeurs.

Il est important de veiller à exposer dans l'entrepôt de données des liens (@href) interprétables par un fournisseur de service distant. Les liens peuvent se trouver à différents niveaux de l'arborescence.

- Au **premier niveau** de la description, en @url de l'élément <ead>.

Exemple :

```

- <ead xsi:schemaLocation="urn:isbn:1-931666-22-9 http://www.archivesportaleurope.net/Portal/profiles/apeEAD.xsd
  http://www.w3.org/1999/xlink http://www.loc.gov/standards/xlink/xlink.xsd" audience="external">
- <eadheader countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b"
  repositoryencoding="iso15511" scriptencoding="iso15924" relatedencoding="MARC21">
  <eadid identifier="FR-AN FRAN IR 017491" countrycode="FR" mainagencycode="FR-AN"
  url="https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/FRAN_IR_017491">FRAN_IR_017491</eadid>

```

- Au niveau composant <c> de la description, dans l'élément <unitid>.

Exemple :

```

- <c level="" id="a0114224631324fziel">
- <did>
  <unittitle>du 22 novembre 1924 au 21 octobre 1925</unittitle>
- <unitid>
  <extptr xlink:href="http://oaipmh.lille.mnesys.fr/ark:/74900/a0114224631324fziel"/>
  1D/1/10
  </unitid>
  <unitdate normal="1924-01-01/1925-12-31" type="inclusive">1924 - 1925</unitdate>
  <physdesc> </physdesc>
</did>

```

- Au niveau images liées <dao>, un lien vers la visionneuse ainsi qu'un lien vers l'image vignette sont attendus.

Exemple :

```

<dao xlink:href="https://www.siv.archives-nationales.culture.gouv.fr/siv/rechercheconsultation/consultation/multimedia
/Galerie.action?mediaParam==?UTF-
8?B?RIJBT18wMTU5XzAxMDUyX0wubXNwI0ZSQU5fMDE1OV8wMTA4Ni9MLm1zcA==?=&udTitle==?UTF-
8?B?UEExBTi1ERS1MQS1UT1VS?=&xpointer=&mmName" xlink:title="Inventaire après décès d'Antoinette de Barats,
femme de Gabriel Nicolas de La Reynie." xlink:role="UNKNOWN"/>
<dao xlink:href="https://www.siv.archives-nationales.culture.gouv.fr/mm/media/download/FRAN_0159_01052_L.msp-
min.jpg" xlink:title="thumbnail"/>

```

### 3.1.2 Transmission d'identifiants fiables

Afin de permettre une chaîne de publication cohérente, l'automatisation des mises à jour, et la génération de liens pérennes sur le portail FranceArchives, il est important de transmettre des identifiants fiables et stables.

Pour garantir la pérennité et l'unicité des identifiants des descriptions encodées en XML-EAD il convient de prêter attention à deux niveaux de granularités :

- L'identifiant EAD : <eadid> qui identifie l'ensemble de l'instrument de recherche.
- L'attribut @id des composants <c> : qui permet d'identifier chaque unité de description. Il est vivement recommandé d'exposer dans l'entrepôt des identifiants générés en lien avec un organisme accrédité. Dans le cas des identifiants ARK il convient de convertir les séparateurs des URI slash (« / ») en tirets (« -- »).

### 3.1.3 Sécuriser l'accès à l'entrepôt

Un service partenaire peut souhaiter filtrer sur IP l'accès à son entrepôt pour des questions de sécurité. Si le service souhaite être fournisseur de données pour le portail FranceArchives tout en filtrant l'accès à son entrepôt, contactez l'équipe FranceArchives pour obtenir les renseignements nécessaires.

### 3.1.4 Paramétrer l'entrepôt

Nous conseillons les paramètres techniques suivants :

- Nombre d'instruments de recherche par requête : cela va dépendre de la taille de vos fichiers, il est raisonnable de commencer par 1.
- Expiration des sessions : minimum 10 minutes. Il peut être intéressant notamment dans un contexte de recette de l'entrepôt d'étendre à 24h afin de permettre un dialogue autour des *restumptionToken* entre fournisseur de données et fournisseur de services.

### 3.1.5 Moissonnage différentiel

Une fois l'entrepôt moissonné par un fournisseur de services, il convient de permettre un moissonnage différentiel des métadonnées présentes dans l'entrepôt à l'aide du paramètre *from* permettant de ne récupérer que les données publiées depuis le dernier moissonnage.

Pour plus d'information consulter la section *SelectiveHarvesting* de la documentation OAI :

<http://www.openarchives.org/OAI/openarchivesprotocol.html#SelectiveHarvesting>

### 3.1.6 Automatisation des suppressions

Il convient de signaler dans l'entrepôt les *Records* qui auraient été supprimés pour permettre l'automatisation des mises à jour. Les réponses à une interrogation *GetRecord* doivent envoyer un *header* contenant un attribut `status="deleted"` sans inclure de *metadata*. Le moissonneur pourra ainsi supprimer le *Record* correspondant via l'identifiant OAI contenu dans le *header*.

Pour plus d'information consulter la section *DeletedRecords* de la documentation OAI :

<http://www.openarchives.org/OAI/openarchivesprotocol.html#DeletedRecords>

## 3.2 Fournisseur de données

Le portail FranceArchives agrège les métadonnées archivistiques des services partenaires et permet leur interrogation conjointe via son moteur de recherche. Les métadonnées archivistiques, quel que soit leur format d'entrée, sont ensuite converties au format apeEAD pour être exposées sur le portail européen.

L'entrepôt a été développé à partir de l'outil open source pyoai (<https://pypi.org/project/pyoai/>).

### 3.2.1 Requêtes et formats supportés

L'entrepôt OAI-PMH de FranceArchives implémente les 6 types de requêtes (verbes) du protocole :

- Identify
- ListMetadataFormats
- GetRecords
- ListIdentifiers
- ListRecords
- ListSets

Le seul format supporté actuellement est le format apeEAD (`metadataPrefix=ape_ead`).

On supporte le moissonnage sélectif des instruments de recherche à l'aide des Sets :

- findingaid : export de tous les instruments de recherche publiés ;
  - findingaid :service :<code\_service> : export de tous les instruments publiés d'un service particulier.
- Ainsi une requête pour obtenir la liste de tous les instruments de recherche publiés de la Vendée prend la forme :

[https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape\\_ead&set=findingaid:service:FRAD085](https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape_ead&set=findingaid:service:FRAD085).

### 3.2.2 Exemples de requêtes

- Identify :

<https://francearchives.fr/oai?verb=Identify>

- ListMetadataFormats :

<https://francearchives.fr/oai?verb=ListMetadataFormats>

- ListSets :

<https://francearchives.fr/oai?verb=ListSets>

- ListIdentifiers avec ou sans filtrage set :

[https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape\\_ead](https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape_ead)

[https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape\\_ead&set=findingaid](https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape_ead&set=findingaid)

[https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape\\_ead&set=findingaid:service:FRAD071](https://francearchives.fr/oai?verb=ListIdentifiers&metadataPrefix=ape_ead&set=findingaid:service:FRAD071)

- ListRecords avec ou sans filtrage set et metadataPrefix :

[https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape\\_ead](https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape_ead)

[https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape\\_ead&set=findingaid:service:FRAD090](https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape_ead&set=findingaid:service:FRAD090)

- GetRecord avec spécification de l'identifiant. La valeur stable\_id correspond à la dernière partie de l'url d'un instrument de recherche.

Ainsi pour <https://francearchives.fr/findingaid/8b4af219170a56f4586f231d05df132b1d3dbfc4> la valeur stable\_id est 8b4af219170a56f4586f231d05df132b1d3dbfc4 ce qui donne la requête suivante [https://francearchives.fr/oai?verb=GetRecord&metadataPrefix=ape\\_ead&identifiant=8b4af219170a56f4586f231d05df132b1d3dbfc4](https://francearchives.fr/oai?verb=GetRecord&metadataPrefix=ape_ead&identifiant=8b4af219170a56f4586f231d05df132b1d3dbfc4).