

Rejoindre le portail

Comment ça marche ?



Vous ne disposez pas d'un entrepôt OAI-PMH ? Tous vos inventaires sont en PDF ? Ou bien encodés dans un EAD imparfait ?

Rassurez-vous, vous n'êtes pas les seuls : ces questions techniques ne doivent pas être un frein à votre participation au portail !

1. Sélectionner ses données.....	2
2. Préparer ses données.....	2
2.1. Gestion des identifiants.....	2
2.2. Les fichiers XML-EAD.....	3
2.2.1 Affichage des éléments EAD sur le portail.....	3
2.2.2 Redirection de l'internaute vers les sites sources.....	3
2.2. Les fichiers PDF.....	5
2.3. Les tableurs.....	5
3. Indexation.....	6
3.1. Quelle valorisation de l'indexation sur le portail ?.....	6
3.2. Alignements sur des référentiels.....	6
4. Transmission des données.....	7
4.1. Envoi direct des fichiers.....	7
4.2. Le protocole OAI-PMH.....	7
4.2.1. Moissonnage XML-EAD.....	8
4.2.2. Moissonnage Dublin Core.....	10
5. Exposition des données du portail.....	10
5.1 Export OAI-PMH du portail.....	11
6. Contact.....	12

1. Sélectionner ses données

Les services contributeurs choisissent eux-mêmes les données qu'ils souhaitent publier sur le portail. Nous n'imposons aucune obligation de publication mais nous vous rappelons l'intérêt, pour le public, de disposer du maximum d'informations provenant de sources validées et conformes à la réglementation sur la communicabilité et la protection des données personnelles. Exemple : Séries anciennes, séries W, état-civil, fonds numérisés, etc.

Nous vous rappelons par ailleurs que le portail est **réservé à la valorisation des archives** et vous remercions par avance de la sélection que vous réaliserez parmi vos fonds : le portail ne valorise pas les bibliothèques, les ouvrages imprimés, ou la presse.

Pour rappel, les données publiées sur FranceArchives sont transformées au format XML-apeEAD (quel que soit leur format d'origine) pour être déposées sur le [portail européen des archives](#) ainsi que sur la [plateforme de données ouvertes](#) du ministère de la Culture en vue de potentielles réutilisations.

Le portail est en mesure d'importer :

- ✓ Des fichiers XML-EAD
- ✓ Des fichiers PDF texte (non image)
- ✓ Des tableurs (fichiers CSV, etc.)



FranceArchives n'héberge pas d'images

Vous pouvez toutefois signaler sur FranceArchives des images numérisées disponibles sur votre site en nous transmettant les métadonnées de ces dernières assorties de liens URL. L'internaute sera ainsi redirigé vers le site internet de votre service pour la consultation de ces images.

2. Préparer ses données

Quel que soit le format utilisé, nous vous suggérons de nous envoyer un échantillon de vos données afin que nous puissions les tester sur le portail et vous faire un premier retour.

2.1. Gestion des identifiants

La chaîne de gestion de fichiers sur le portail est fondée sur la récupération d'un identifiant fourni par votre service. L'identifiant détermine l'exécution des tâches suivantes au sein de FranceArchives :

- Premier chargement de l'inventaire,
- Attribution d'un identifiant pérenne sur le portail,
- Mise(s) à jour de l'inventaire.

Le type d'identifiant récupéré par le portail dépend du protocole d'import :

Protocole	Identifiant utilisé	Formats importés
OAI-PMH	Identifiant EAD : <eadid>	XML-EAD
HTTP (zip)	Nom du fichier	XML-EAD, CSV, PDF


Afin d'harmoniser la gestion des fichiers XML-EAD indépendamment du protocole d'import, il est nécessaire que pour un instrument de recherche, le nom de fichier et l'identifiant EAD (<eadid>) soient identiques.

Identifiants EAD et noms de fichiers doivent être formés de la manière suivante :

- Identifiant EAD : [Code-institution](#)_code_inventaire
- Nom de fichier : [Code-institution](#)_code_inventaire.[extension](#)


Le code institution permet de relier l'instrument de recherche au service responsable de ce dernier, répertorié dans l'annuaire du portail. Il est souvent construit à partir des normes [ISO/FDIS 15511:2011 Information et documentation – Identifiant international normalisé pour les bibliothèques et les organismes apparentés \(ISIL\)](#), et [ISDIAH](#) : Norme internationale pour la description des institutions de conservation des archives.

- Exemples de code institution : FRAN, FRAD008, FRAC06406, FRAC34172, etc.
- Exemple d'identifiant EAD : <eadid>FRSHD_INV_GR28P12</eadid>, <eadid>FRAPHP075_000568</eadid>

	<p>Veillez à ne pas attribuer le même identifiant à deux instruments de recherche différents, et à bien conserver le même identifiant entre deux mises à jour.</p> <p>Veillez à n'utiliser ni espaces, ni caractères spéciaux dans les noms de fichiers et identifiants EAD.</p>
--	--

2.2. Les fichiers XML-EAD

Pour pouvoir être importé sur le portail, un instrument de recherche encodé en XML doit être bien formé : c'est-à-dire qu'il doit respecter les règles syntaxiques de l'XML, mais pas nécessairement être valide du point de vue de la DTD EAD.

	<p>Vous pouvez tester vos fichiers très simplement, en les ouvrant sur un navigateur : s'il n'apparaît pas de message d'erreur, votre fichier est bien formé.</p>
---	---

2.2.1 Affichage des éléments EAD sur le portail

Pour mieux comprendre l'affichage des éléments EAD sur le portail, nous vous renvoyons à la consultation du [tableau d'affichage des éléments](#). Voir aussi : [3. Indexation](#)

2.2.2 Redirection de l'internaute vers les sites sources

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.

Dans un fichier XML-EAD, un attribut HREF contient l'adresse d'une ressource au sein d'un élément. Dans la mesure où la ressource est distante, la valeur attendue de l'attribut HREF sur le portail FranceArchives est un URL (Uniform Resource Locator = Localisateur univoque de ressource) ou un URI (Uniform Resource Identifier).

La restitution des URL/URI peut se traduire à plusieurs endroits de l'arborescence XML-EAD. Voici plusieurs exemples de restitution (potentiellement cumulables) :

- 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="a0114224631324fzieI">
- <did>
  <unittitle>du 22 novembre 1924 au 21 octobre 1925</unittitle>
- <unitid>
  <extptr xlink:href="http://oai.pmh.lille.mnesys.fr/ark:/74900/a0114224631324fzieI"/>
  1D/1/10
  </unitid>
  <unitdate normal="1924-01-01/1925-12-31" type="inclusive">1924 - 1925</unitdate>
  <physdesc> </physdesc>
</did>
```

- Pour restituer des images, au sein des balises <dao> ou <daogrp>. L'exemple ci-dessous comprend un lien vers la visionneuse, ainsi qu'un lien vers la vignette. Le lien vers la vignette n'est pas à négliger car il permet au portail d'afficher cette dernière sur la page de résultats de recherche et sur celle du composant, augmentant de ce fait le signalement à l'internaute d'une numérisation à consulter.

Exemple :

```
<dao xlink:href="https://www.siv.archives-nationales.culture.gouv.fr/siv/rechercheconsultation/consultation/multimedia
/Galerie.action?mediaParam==?UTF-
8?B?RIJBT18wMTU5XzAxMDUyX0wubXNwI0ZSQ5fMDE1OV8wMTA4Ni9MLm1zcA==?=&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"/>
```

Dans le cas où la valeur de l'attribut HREF ne serait pas un URL mais un chemin relatif, le lien sera inopérant en dehors de son environnement spécifique, et donc sur FranceArchives. Afin de pallier en partie ce problème, nous paramétrons si possible une requête automatique vers le moteur de recherche du site source (cote ou eadid). Ainsi l'internaute n'a pas à ressaisir entièrement sa recherche, un certain nombre de résultats lui sont proposés sur le site source.

2.2. Les fichiers PDF

Sur le portail, les fichiers PDF sont indexés en plein texte, c'est-à-dire que si une expression recherchée par l'internaute est présente dans le corps d'un document PDF, ce dernier apparaîtra dans les résultats de recherche. Il est conseillé toutefois d'ajouter des métadonnées à ces fichiers non structurés :

- Afin qu'ils soient plus efficacement indexés par le moteur de recherche,
- Afin de guider l'internaute jusqu'au document.

En effet, ces métadonnées constituent une interface incontournable pour l'internaute : elles représentent le contenu du document sur la page de résultat de recherche, puis sur la notice intermédiaire proposant l'ouverture du fichier PDF.

Des métadonnées pertinentes permettent un meilleur référencement du document et augmentent ses chances d'être consulté.

- ➔ Si vos fichiers PDF sont imbriqués dans un « squelette XML-EAD » ils peuvent être importés sur FranceArchives avec le fichier XML qui les relie.
- ➔ Si vos fichiers PDF ne possèdent pas de métadonnées, nous vous proposons de remplir un fichier spécifique :

[Télécharger le fichier de métadonnées au format csv¹,](#)

[Télécharger le fichier de métadonnées au format xlsx.](#)

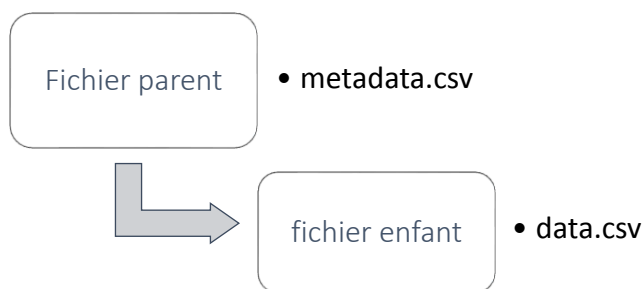
Il est tout à fait possible de remplir ce tableau de façon très sommaire dans un premier temps (titre, dates extrêmes) et de nous renvoyer des versions enrichies par la suite.

Voir aussi : [3. Indexation](#)

2.3. Les tableurs

L'import de tableurs CSV permet notamment la diffusion sur FranceArchives d'informations exportées depuis une base de données relationnelle. Une description à deux niveaux est permise par la gestion de deux fichiers différents :

¹ Pour ces fichiers le séparateur utilisé est la virgule « , ».



[Télécharger le fichier data.csv](#) ²

[Télécharger le fichier metadata.csv](#)

3. Indexation

Nous vous incitons vivement à indexer les instruments de recherche de vos services, en vous référant à la note d'information [DITN/RES/2007/008](#).

Les éléments d'indexation privilégiés sur le portail sont contenus :

- À l'intérieur de l'élément Vedettes et accès contrôlés <controlaccess> des fichiers XML-EAD,
- Au sein des fichiers metadata.csv pour les fichiers PDF (voir : [2.2. Les fichiers PDF](#)).

3.1. Quelle valorisation de l'indexation sur le portail ?

L'indexation des fichiers XML et PDF permet d'améliorer le référencement des instruments de recherche sur le moteur de recherche du portail. Par ailleurs, l'auto-complétion proposée sur le moteur s'appuie uniquement sur les termes d'indexation des instruments de recherche (XML et PDF). De ce fait, une recherche utilisant l'auto-complétion remontera uniquement les instruments de recherche indexés avec l'expression sélectionnée par l'internaute.

3.2. Alignements sur des référentiels

Un enrichissement des données collectées sur le portail est prévu à travers la réalisation d'alignements des données d'indexation vers des référentiels externes. Ce type d'opération permet de « désambiguïser » les termes d'indexation en leur associant des concepts univoques, compréhensibles par les humains comme par les machines. Ces alignements auront pour effet d'homogénéiser les données d'indexation produites par divers services d'archives, mais également d'améliorer leur visibilité sur les moteurs de recherche généralistes et, de ce fait, d'améliorer la visibilité des ressources associées à ces concepts.

² Pour ces fichiers CSV le séparateur utilisé est la virgule « , ».

Les noms géographiques (éléments <geogname>) ainsi que les noms de personnes physiques (éléments <persname>) seront traités prioritairement. Le travail lié à l'alignement de ces éléments d'indexation sur les référentiels [GeoNames](#) et [data.bnf](#) passera en production courant 2019.

Les données enrichies seront par la suite reversées à votre service, afin que vous puissiez bénéficier du fruit de ces traitements.

4. Transmission des données

Le processus de signature des conventions, qui peut parfois prendre plusieurs mois, peut avoir lieu en parallèle du processus de test et d'import des données. N'hésitez pas à nous faire parvenir vos données avant même la signature de la convention : un échantillon représentatif pour test puis l'ensemble des données à publier, nous serons ainsi en mesure de diffuser vos données dès la signature de la convention. Ces données ne seront dans tous les cas mises en ligne qu'après signature de la convention.

Pour les mises à jour et/ou enrichissements, un calendrier est défini avec chaque service contributeur. Nous préconisons toutefois de réaliser une **mise à jour de vos données tous les 6 mois**.

4.1. Envoi direct des fichiers

Vous pouvez nous transmettre vos données par mail au sein d'une archive zip, ou par le biais d'un serveur de fichier (type [Framasoft](#), [Zephyrin](#), [Wetransfer](#), etc.).

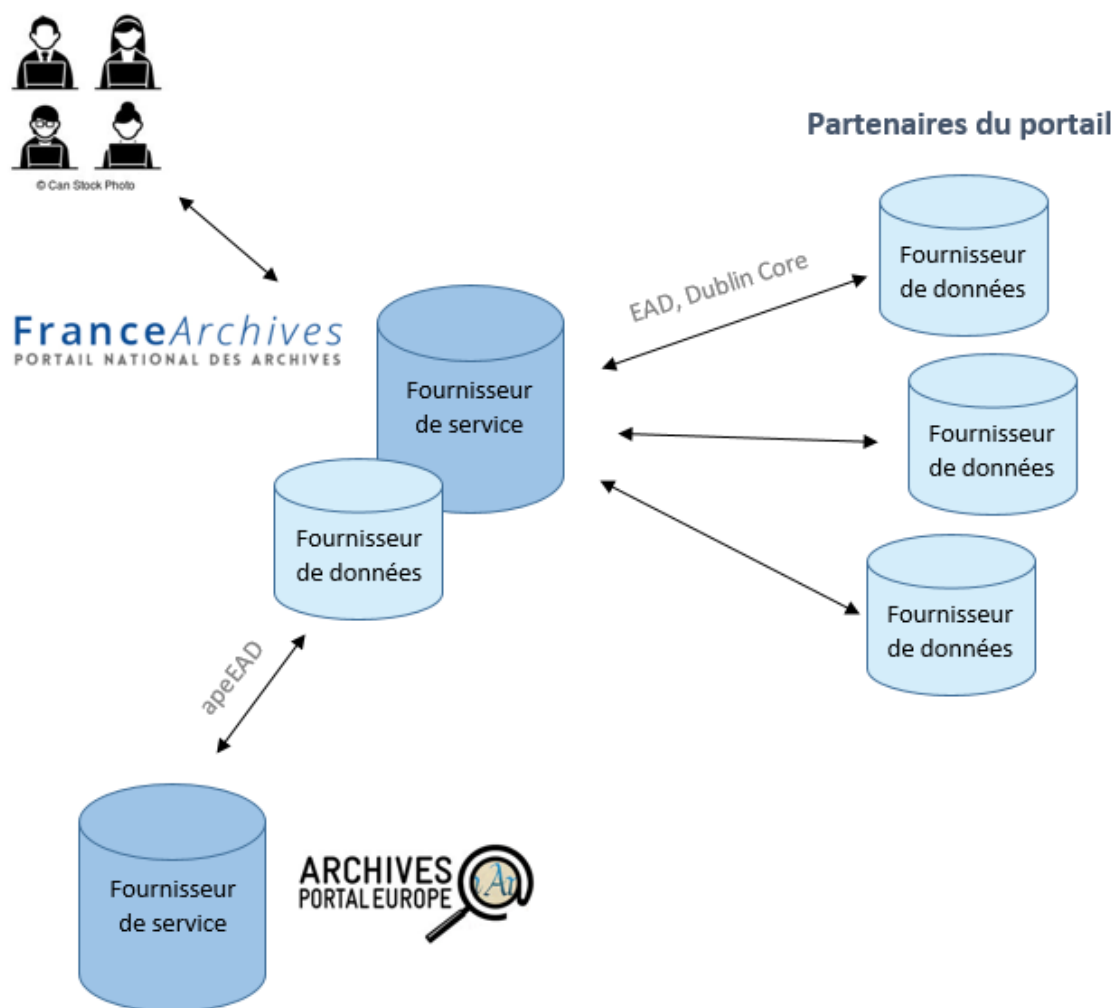
4.2. Le protocole OAI-PMH

Si vous avez un entrepôt OAI-PMH, il peut être moissonné par FranceArchives. L' [OAI-PMH](#), permet d'exposer et d'échanger des métadonnées sur le Web. Il fonctionne sur la base du HTTP, protocole standard du web, enrichi de six verbes d'interrogations spécifiques :

- Identify
- ListMetadataFormats
- GetRecord
- ListIdentifiers
- ListRecords
- ListSets

Un entrepôt OAI propose deux niveaux de granularité : le *Set* et le *Record*. Le *Set* permet de constituer un ensemble cohérent de ressources déterminé par le fournisseur de données. Le *Record* désigne une ressource et ses métadonnées associées, par exemple pour l'XML-EAD le *Record* désigne ainsi l'instrument de recherche encodé. Le fournisseur de données doit indiquer au fournisseur de service le *set* à moissonner.

Le schéma ci-dessous représente la chaîne de transmission des données sur FranceArchives :



Le protocole OAI-PMH permet l'échange de métadonnées XML selon différents formats. Les formats supportés à l'import par le portail FranceArchives sont les suivants :

- EAD (metadataPrefix=ead ou metadataPrefix=oai_ead)
- Dublin Core (metadataPrefix=oai_dc)
- Nomina (metadataPrefix=nomina)

4.2.1. Moissonnage XML-EAD

Chaque instrument de recherche encodé en XML-EAD correspond à un *record*. Chaque élément <record> comprend un élément <metadata> contenant lui-même l'élément <ead> correspondant à l'instrument de recherche moissonné.

Exemple :


```

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2016-12-06T10:10:23Z</responseDate>
  <request verb="ListRecords" metadataPrefix="ead" set="francearchives"
    >https://recherche-archives.doubs.fr/oai_pmh.cgi</request>
  <ListRecords>
    <record>
      <header>
        <identifier>25993/a011369750208WU2TRM</identifier>
        <datestamp>2016-12-05T15:56:37Z</datestamp>
        <setSpec>intendance_de_franche-comte</setSpec>
      </header>
      <metadata>
        <ead audience="external" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:isbn:1-931666-22-9" xmlns:xlink="http://www.w3.org/1999/xlink"
  xsi:schemaLocation="urn:isbn:1-931666-22-9 http://www.loc.gov/ead/ead.xsd
http://www.w3.org/1999/xlink http://www.loc.gov/standards/xlink/xlink.xsd">
          <eadheader findaidstatus="edited-full-draft">
            <eadid>FRAD025_000979_62J</eadid>
            <filedesc>
              <titlestmt>
                <titleproper>62J Ordre et syndicat des architectes</titleproper>
                <subtitle/>
              </titlestmt>
            </eadheader>
            <archdesc level="fonds">
              (...)
            </archdesc>
          </ead>
        </metadata>
      </record>
    </ListRecords>
  </OAI-PMH>

```

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 ou 2.
- Expiration des sessions : 10 minutes.

Nous vous conseillons dans un premier temps de commencer par charger une dizaine d'inventaires et de s'assurer du bon fonctionnement de l'entrepôt par quelque test.

4.2.2. Moissonnage Dublin Core

Le Dublin Core Metadata Element Set, aussi désigné « Dublin Core Simple » comprend quinze propriétés de base. Il est notamment utilisé par les services d'archives pour l'exposition de bases images. Le portail moissonne des données au format Dublin Core selon un profil spécifique (voir : les [recommandations oai_dc FranceArchives](#)).

Exemple :

```
<OAI-PMH
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2018-07-20T11:59:59+02:00</responseDate>
  <request metadataPrefix="oai_dc"> http://archives.meuse.fr/oaipmh/repository/ </request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:archives.meuse.fr:a011500569427JPvaBc</identifier>
        <datestamp>2018-06-29</datestamp>
        <setName>État civil</setName>
        <setSpec>FRAD055_EC</setSpec>
      </header>
      <metadata>
        <oai_dc:dc
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Naissances, Mariages, Décès (1883-1892)</dc:title>

          <dc:date>1883</dc:date>
          <dc:subject>Naissances ; Mariages ; Décès</dc:subject>
          <dc:coverage>Les Marats (Meuse, France)</dc:coverage>
          <dc:source>2 E 326 (11) </dc:source>
          <dc:identifier>http://archives.meuse.fr/ark:/52669/a011500569427JPvaB
        </dc:identifier>

          <dc:relation> http://archives.meuse.fr/ark:/52669/a011500569427JPvaBc
        </dc:relation>
        </oai_dc:dc>
      </metadata>
    </record>
```

5. Exposition des données du portail

Les données transmises par les partenaires du portail (XML-EAD, Dublin Core, PDF, CSV) sont par la suite converties sur le portail FranceArchives vers le format [XML-apeEAD](#), dénominateur commun européen entre les différentes versions de l'EAD.

Par ailleurs le portail expose les instruments de recherche en RDF selon le modèle de données d'Europeana (EDM Data Model). Pour générer le RDF, ajouter /rdf.xml ou /rdf.n3 ou /rdf.nt à une URI du portail. Exemple :

<https://francearchives.fr/fr/findingaid/79f06b8aa81de090a15d88f5e1c6154f6ed91f4e/rdf.xml>

5.1 Export OAI-PMH du portail

FranceArchives a un entrepôt qui expose les données reçues des services partenaires, il est moissonné par le portail européen. Les 6 types de requêtes du protocole OAI-PMH sont implémentées :

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

Le seul format supporté est le format apeEAD : `metadataPrefix=ape_ead`

- Moissonnage sélectif :

Des sets sont constitués par services d'archives pour permettre un moissonnage sélectif.

- *findingaid* : export de tous les instruments de recherche publiés ;
- *findingaid:service:<code_service>* : export de tous les instruments publiés d'un service particulier.

Voici un exemple de requête pour obtenir la liste de de tous les instruments de recherche de la Vendée publiés sur le portail FranceArchives :

https://francearchives.fr/oai?verb=ListRecords&metadataPrefix=ape_ead&set=findingaid:service:FRAD085

- 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&set=findingaid

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&set=findingaid:service:FRAD090

- *GetRecord* avec spécification de l'*identifiant* :

La valeur *stable_id* correspond à l'identifiant pérenne de l'instrument de recherche correspondant à la dernière partie de l'URL. Par exemple 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

- Format des enregistrements *record* des réponses OAI-PMH

Pour les requêtes *GetRecord* et *ListRecords*, la réponse OAI-PMH contient deux balises à l'intérieur de la (ou des) balise(s) <record> :

- la balise <header>, qui contient l'*identifiant* de l'enregistrement ainsi que sa date de modification ;
- la balise <metadata> qui contient les données de l'enregistrement dont le format dépend du type d'objet de la requête.

6. Contact

Pour toute question d'ordre technique liée au portail FranceArchives, n'hésitez pas à nous contacter : francearchives.siaf@culture.gouv.fr.