OAI-PMH

Table des matières

1 Intro

On peut glaner sur HAL, via une API web :

- HAL
- thèse
- image

2 oai_browser

Mais on peut aussi moissoner via OAI-PMH :

```
# apt-get install libhttp-oai-perl
# man oai_browser
$ oai_browser https://api.archives-ouvertes.fr/oai/hal/

>4 ListMetadataFormats
-------------------
Enter an optional identifier>
...
metadataPrefix => oai_dc
metadataPrefix => oai_dcterms
metadataPrefix => xml-tei
```

Pour lister les identifiants : ListIdentifiers

```
>3 ListIdentifiers
---------------
Enter an optional resumptionToken>
Enter an optional from period (yyyy-mm-dd)>
Enter an optional until period (yyyy-mm-dd)>
Enter an optional set ([A-Z0-9_]+)> collection:ETHNO
Enter the metadataPrefix to use> oai_dc
identifier => oai:HAL:halshs-01470887v1 / 2017-02-18
...
identifier => oai:HAL:halshs-01380221v1 / 2016-10-13
^C
```

Pour récupérer les méta-données : GetRecord

```
>1 GetRecord
---------
Enter the identifier to request> oai:hal:tel-00766707v2
Enter the metadataPrefix to use> oai_dc
...
```

```
>1 GetRecord
---------
Enter the identifier to request> oai:hal:tel-00766707v2
Enter the metadataPrefix to use> oai_dcterms
...
  <dcterms:identifier>tel-00766707</dcterms:identifier>
  <dcterms:identifier>https://tel.archives-ouvertes.fr/tel-00766707</dcterms:identifier>
  <dcterms:identifier>https://tel.archives-ouvertes.fr/tel-00766707v2/document</dcterms:identifier>
  <dcterms:identifier>https://tel.archives-ouvertes.fr/tel-00766707/file/ThA_se_CHABROL_Fanny_-_HAL.pdf</dcterm
...
```

La première URL pointe sur la fiche web descriptive dans HAL, les suivante sur le document dans divers formats. autre exemple :

```
>1 GetRecord
---------
Enter the identifier to request> oai:hal:medihal-00508315v1
Enter the metadataPrefix to use> oai_dcterms
...
  <dcterms:identifier>medihal-00508315</dcterms:identifier>
  <dcterms:identifier>https://medihal.archives-ouvertes.fr/medihal-00508315</dcterms:identifier>
  <dcterms:identifier>https://medihal.archives-ouvertes.fr/medihal-00508315/image</dcterms:identifier>
  <dcterms:identifier>https://medihal.archives-ouvertes.fr/medihal-00508315/file/catherineperrier_Jouet_V2.jpg<
  <dcterms:identifier>https://medihal.archives-ouvertes.fr/medihal-00508315/file/catherineperrier_Jouet_V2.tiff
...
```

Moissoner veut bien dire ce que ça veut dire : on a la date et un seul critère de recherche (avec seulement une description accessible) :

```
$ echo 6 | oai_browser https://api.archives-ouvertes.fr/oai/hal/ > /tmp/sets.txt

$ grep -A1 ETHNO /tmp/sets.txt
setSpec => collection:ETHNO
setName => Productions de l'ethnologie
--
setSpec => collection:AO-ETHNO
setName => Archives ouvertes de l'Ethnologie
```

When a setSpec is used as an argument, the response must include :

- the records corresponding to the metadataPrefix argument, or headers thereof in the case of deleted records, available from those items in the set specified by the setSpec ;
- the records corresponding to the metadataPrefix argument, or headers thereof in the case of deleted records, available from those items in sets that are descendant from the specified set.

3 HTTP : :OAI

Idem via un script :

- ListIdentifiers
- GetRecord