# **OAIP-MH Service: een kort overzicht van de mogelijkheden en werking**

1. **Inleiding**

Collective Access voorziet in een ‘OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting’ implementatie.  
Alle mogelijke requests (of verbs (zie verder)) zijn daarbij geïmplementeerd.

Om uw data beschikbaar te maken via OAI-PMH wordt, per gevraagd XML-formaat (Simple Dublin Core, Qualified Dublin Core, Ebucore of LIDO), een export mapping aangemaakt (een speciale Excel spreadsheet). U vindt deze mappings in de CA-interface terug onder BEHEREN -> Exporters.   
Opm 1: mappings kunnen via de CA-interface enkel opgeladen of verwijderd worden.   
Opm 2: U ontvangt uiteraard een kopie van de Excel spreadsheets.  
Opm 3: Momenteel is enkel oai\_dcterms en **ebucore** beschikbaar.

De verdere definitie dient door ons te gebeuren in een configuratie bestand.

Hoe spreken we deze service nu aan?

- Hoe zien de verschillende requests eruit?

- Hoe bepaalt u welke informatie beschikbaar gesteld wordt?

1. **Hoe zien de verschillende requests eruit?**

Een request wordt opgebouwd uit:

* de basisurl;
* gevolgd door een ‘?’,
* gevolgd door verb= ,
* (gevolgd door 0 of meerdere parameters, gescheiden door ‘&’).
  1. **De BasisUrl:**

[**https://www.hetvirtueleland.be/ca\_cag/service.php/OAI/ebucore\_cag\_film/request**](https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request)

Hierin is ‘**ebucore\_cag\_film**’ een naam die wij gekozen hebben in het configuratiebestand.

Wenst U hier een andere benaming te gebruiken, laat het ons weten.

* 1. **De Verbs**

Het OAI-PMH protocol voorziet in volgende verbs (= een type van request).

* + 1. ***Identify***

**Doel**: informatie opvragen over uw ‘repository’, zoals de naam, email, etc…

**Argumenten**: <geen>

**Voorbeeld**:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=Identify>

* + 1. ***ListMetadataFormats***

**Doel**: informatie opvragen over welke formaten voor de metadata beschikbaar zijn in uw repository. (Voor uw basisurl : **ebucore**)

**Voorbeeld**:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListMetadataFormats>

**Argumenten**: <geen>

* + 1. ***ListSets***

**Doel**: informatie opvragen over de beschikbare collecties. Via deze optie wordt ‘selective harvesting’ mogelijk (zie toelichting)

**Voorbeeld**:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListSets>

**Argumenten**: <geen>

**Toelichting**:  
Wat een set is, wordt bepaald in het CA configuratiebestand. Momenteel zijn de sets verbonden met de **object types**, zonder enige beperking.

* + 1. ***ListIdentifiers***

**Doel**: opvragen van de ‘headers’ van de records, i.p.v. het volledige record. (Verkorte vorm van ListRecords)

**Voorbeeld**:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListIdentifiers>

**Argumenten**: from, until, set, resumptionToken

Zonder ‘Argumenten’ krijgen we alle objecten die **publiek toegankelijk** zijn terug. Voor het gebruik van de Argumenten: zie verder.

* + 1. ***ListRecords***

**Doel**: harvesten van de inhoud van de repository in het gewenste formaat.

**Voorbeeld**:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords>

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&metadataPrefix=ebucore>

Zonder opgave van het argument **metadataPrefix** krijgt U het default metadata formaat terug.  
De bovenstaande URLs geven dus beiden hetzelfde resultaat.

Wenst U echter een ander formaat dan de default, dan is het argument **metadataPrefix=** verplicht, zoals bv voor **qdc** (qualified Dublin Core).

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&metadataPrefix=oai_dcterms>

**Argumenten**: from, until, set, resumptionToken en metadataPrefix (noodzakelijk voor een ander dan het default format)

* + 1. ***GetRecord***

**Doel**: opvragen van de metadata van een individueel record.

**Voorbeeld:**

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=GetRecord&metadataPrefix=ebucore&identifier=oai:oai.cag:38790>

wat gelijk is aan

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=GetRecord&identifier=oai:oai.cag:38790>

(Voor qualified Dublin Core wordt dat:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=GetRecord&metadataPrefix=oai_dcterms&identifier=oai:oai.cag:38790>)

**Argumenten:** identifier (verplicht), metadataPrefix (noodzakelijk voor een ander dan het default formaat)

Waar vinden we de informatie voor het argument **identifier=** ?

Deze informatie vindt U terug in het resultaat van het ListIdentifiers-verb.

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8" standalone="yes"?>  <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>2020-07-17T07:52:22Z</responseDate>  <request verb="ListIdentifiers" metadataPrefix="ebucore">http://www.hetvirtueleland.be/ca\_cag\_17\_test/service.php/OAI/ebucore\_cag\_film</request>  <ListIdentifiers>  <header>  <identifier>oai:oai.cag:1</identifier>  <datestamp>2020-02-19T10:17:36Z</datestamp>  </header>  <header>  <identifier>oai:oai.cag:4</identifier>  <datestamp>2019-09-17T14:32:49Z</datestamp>  </header>  <header>  <identifier>oai:oai.cag:7</identifier>  <datestamp>2019-09-17T14:32:49Z</datestamp>  </header>  <header>  <identifier>oai:oai.cag:10</identifier>  <datestamp>2019-09-17T14:34:57Z</datestamp>  </header>  … |

De url:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&metadataPrefix=ebucore>

geeft alle records terug, zonder rekening te houden met de sets, die verwijzen naar het object\_type.

Willen we alle records, behorende tot de set films, dan wordt de url:

#### <https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&metadataPrefix=ebucore&set=34668>

* 1. **Parameters**

Tot slot nog een woordje uitleg bij de parameters en hun gebruik.

* + 1. ***metadataPrefix***

De parameter waarmee u aangeeft welk formaat gebruikt moet worden voor het antwoord.

Voor onze basisurl is enkel **oai\_dcterms**  en **Ebucore** als format voorzien.

Voorbeelden: zie 2.2.5

* + 1. ***set***

De parameter waarmee U (bijkomend) aangeeft uit welke set (= object\_type) U records wil harvesten.  
De beschikbare sets kan U opvragen m.b.v. het **ListSets** verb. (zie 2.2.3)  
Zonder opgave van ‘**set=34668**’ krijgt U alle publiek toegankelijke records als antwoord (los van de set waartoe ze behoren).  
Met opgave van ‘**set=34668**’ krijgt U enkel de publiek toegankelijke records in set=34688 – Films.

Set staat hierbij voor een object\_type.   
Opdat een object\_type zichtbaar is met **verb=ListSets**, moet ze publiek toegankelijk zijn (***en*** moet er minstens één record van deze collectie publiek toegankelijk zijn – zie 2.2.3)

* + 1. ***from en until***

Deze parameters maken ‘verdere **selective harvesting**’ mogelijk.   
Met **from** geeft U een begindatum (en uur) van harvesten op, met **until** geeft U een einddatum (en uur van harvesten op.  
Alle records met een create, update of delete vlag in de betrokken tijdsperiode worden geharvest.

Voorbeeld:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListIdentifiers&from=2018-09-11&until=2018-09-13>

Opgelet: de CA OAI-PMH implementatie laat niet toe dat begindatum gelijk is aan einddatum.

* + 1. ***identifier***

Met deze parameter geeft u de unieke identificatie op van het record dat U wenst te harvesten.  
Deze identifiers kan U opvragen met het ListIdentifiers – verb (zie 2.2.4)

* + 1. ***resumptionToken***

In het configuratie-bestand van OAI wordt eveneens bepaald hoeveel records er per request verstuurd worden (maxRecordsPerRequest).   
(Deze variabele staat tijdelijk maar op 20, juist om deze parameter te kunnen demonstreren. Een normale waarde is 50 of 100)

Als U een request stuur naar de OAI-PMH, dan wordt de omvang van het antwoord berekend (totaal aantal records dat verzonden zou moeten worden (completeListSize)). Dit aantal wordt vervolgens vergeleken met de maxRecordsPerRequest.   
Als er meer records verzonden moeten worden dan door maxRecordsPerRequest bepaald, wordt onderaan de XML van de eerste batch volgende regel toegevoegd (voorbeeld):

|  |
| --- |
| <resumptionToken expirationDate="2018-10-16T13:21:55Z" completeListSize="32" cursor="0">dda2f22881ea8bc318a10b260187d795</resumptionToken> |

Uit deze regel kunnen we het volgende afleiden:

* **completeListSize=”32”**: het volledige antwoord omvat 32 records
* **expirationDate="2018-10-16T13:21:55Z":**  de geldigheidsduur van deze token (maar CA maakt hier geen gebruik van)

Vermits **maxRecordsPerRequest = 20**, bevat dit antwoord 20 van 32 records.   
Er kunnen dus nog 12 records bijkomend geharvest worden.  
Daarvoor hebben we de string dda2f22881ea8bc318a10b260187d795 nodig.

Voorbeeld:

Voor de eerste 20 records, in formaat ebucore:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&metadataPrefix=ebucore>

Voor de resterende records, in formaat ebucore:

<https://www.hetvirtueleland.be/ca_cag/service.php/OAI/ebucore_cag_film/request?verb=ListRecords&resumptionToken=dda2f22881ea8bc318a10b260187d795>

Merk op:   
als U gebruik maakt van een resumptionToken is het NIET meer nodig de metadataPrefix (= formaat) van het antwoord te bepalen.