

Identifiers	baseurl_oai
oai:cn.pdg.it:ca.artsch527	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch528	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch529	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch530	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch531	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch532	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch533	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch534	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch535	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch536	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch537	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch538	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch539	http://www.biblioteca.unige.it
oai:cn.pdg.it:ca.artsch540	http://www.biblioteca.unige.it

Jun 05, 2021

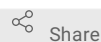
How to create baseurl for OAI-PMH harvesting with OpenRefine

Forked from [Creare baseurl per harvesting di repository OAI con OpenRefine](#)

Andrea Marchitelli¹, Carlo Bianchini², Alessandra Moi¹

¹Associazione Italiana Biblioteche (AIB); ²Università di Pavia

1 Works for me



Share

dx.doi.org/10.17504/protocols.io.bvi9n4h6

wikidata and libraries

Andrea Marchitelli

ABSTRACT

This protocol describes the method for creating baseurl for harvesting an OAI-PMH repository from a list of identifiers.

EXTERNAL LINK

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

DOI

dx.doi.org/10.17504/protocols.io.bvi9n4h6

EXTERNAL LINK

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

PROTOCOL CITATION

Andrea Marchitelli, Carlo Bianchini, Alessandra Moi 2021. How to create baseurl for OAI-PMH harvesting with OpenRefine. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.bvi9n4h6>

FORK NOTE

FORK FROM

Forked from [Creare baseurl per harvesting di repository OAI con OpenRefine](#), Andrea Marchitelli

KEYWORDS

oai-pmh, OpenRefine, data

LICENSE

This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

IMAGE ATTRIBUTION

Le immagini sono realizzate personalmente dall'Autore

CREATED

Jun 05, 2021

LAST MODIFIED

Jun 05, 2021

PROTOCOL INTEGER ID

50497

MATERIALS TEXT

OpenRefine

[source](#) by Freebase, then Google, now open source community

BEFORE STARTING

Installare l'applicativo Openrefine

- 1 Open the OpenRefine project containing the list of OAI identifiers prepared for harvesting. The list can be prepared by following the directions in the linked protocol How to import OAI-PMH identifiers into OpenRefine.



How to import OAI-PMH identifiers in OpenRefine
by **Andrea Marchitelli**

PREVIEW

RUN



- 1.1 Open OpenRefine and choose "Create project".

OpenRefine

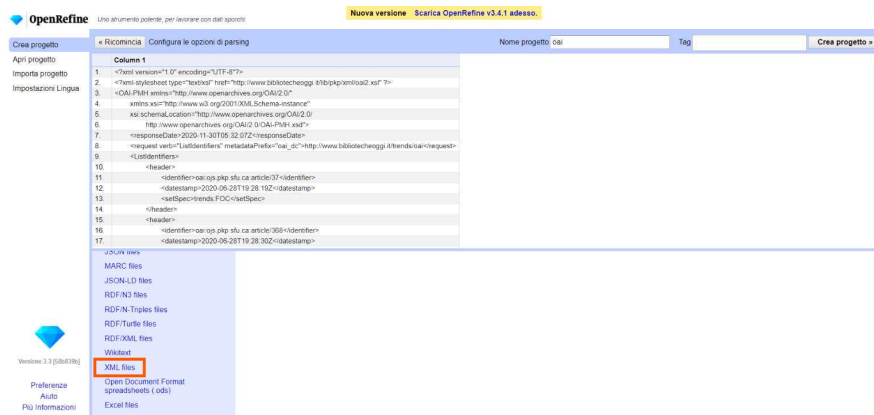
[source](#) by Freebase, then Google, now open source community

- 1.2 Select "Web Addresses (URLs)" as the data source for the new project and click "Next". Enter the OAI-PMH baseurl of the identified target, choose the **ListIdentifiers** verb and the appropriate metadataPrefix.
E.g. http://www.bibliotecheoggi.it/trends/oai?verb=ListIdentifiers&metadataPrefix=oai_dc

The OAI-PMH protocol allows per-date and per-set limiters to be set to expose partial datasets and allow incremental collections.

Es. <http://an.oa.org/OAI-script?verb=ListIdentifiers&from=1998-01-15&metadataPrefix=oldArXiv&set=physics:hep>

- 1.3 Select "XML Files" in the "Parse Data As" section.



Screenshot of import options in OpenRefine

- 1.4 In the preview area of the file to be imported, locate the first line containing the element and bring the mouse pointer to highlight it.
Then click on it.



Highlight the identifier

1.5

At that point you will have obtained the list of OAI-PMH identifiers, one per line, ready for further processing.

Creare le URL per l'harvesting

- 2 Click on the identifier column present in the project and choose the option Edit Column > Add Column based on this column.

2.1

In the dialog with the new column settings enter baseurl_oai as the column name and a GREL expression that concatenates the journal's baseurl_oai, the GetRecord verb, and the OAI identifier already in the files.

E.g.. 'http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier='+value

Be careful to keep superscripts (') at the beginning and end of the URL string

Test del risultato

3



At this point confirming the operation you will obtain a new column called baseurl_oai containing the URLs for the GetRecord of each article.

97 righe

Vista a: **righe** elementi Mostra: 5 10 25 50 righe

Tutti	identificher	baseurl_oai
1.	oai:ijs.pkp.sfu.ca:article/37	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/37
2.	oai:ijs.pkp.sfu.ca:article/368	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/368
3.	oai:ijs.pkp.sfu.ca:article/587	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/587
4.	oai:ijs.pkp.sfu.ca:article/718	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/718
5.	oai:ijs.pkp.sfu.ca:article/834	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/834
6.	oai:ijs.pkp.sfu.ca:article/1004	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/1004
7.	oai:ijs.pkp.sfu.ca:article/384	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/384
8.	oai:ijs.pkp.sfu.ca:article/515	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/515
9.	oai:ijs.pkp.sfu.ca:article/627	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/627
10.	oai:ijs.pkp.sfu.ca:article/776	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ijs.pkp.sfu.ca:article/776

4 Test the result obtained by clicking on a couple of links and verifying that the corresponding page opens correctly