Mostra: 5 10 25 50 righe	
identifier	baseurl_oai
oai:ajs.pkp.sfu.ca:article/37	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/368	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/587	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/718	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/834	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/1004	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/384	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/515	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/627	http://www.bibliotecheoggi.it
oai:ajs.pkp.sfu.ca:article/776	http://www.bibliotecheoggi.it

NOV 30, 2020

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.bp7pmrmn

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

Protocol Citation: Andrea Marchitelli 2020. Creare baseurl per harvesting di repository OAI con OpenRefine. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.bp7pmrmn>

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

Protocol status: Working
 We use this protocol and it's working

Created: Nov 30, 2020

Last Modified: Nov 30, 2020

PROTOCOL integer ID:
 45007

🌐 Creare baseurl per harvesting di repository OAI con OpenRefine

Andrea Marchitelli¹

¹Associazione Italiana Biblioteche (AIB)

wikidata and libraries

Andrea Marchitelli

ABSTRACT

Il protocollo descrive il metodo per creare le baseurl per l'harvesting di un repository OAI-PMH a partire da una lista di identificatori.

IMAGE ATTRIBUTION

Le immagini sono realizzate personalmente dall'Autore

MATERIALS

Software

OpenRefine

NAME

Freebase, then Google, now open source community

DEVELOPER

<https://github.com/OpenRefine/OpenRefine>

SOURCE LINK

BEFORE START INSTRUCTIONS

Installare l'applicativo Openrefine

Keywords: oai-pmh,
OpenRefine, data

- 1 Aprire il progetto OpeRefine contenente la lista degli identificatori OAI predisposta per l'harvesting.
La lista può essere predisposta seguendo le indicazioni del protocollo collegato *Come importare identificatori OAI-PMH in OpenRefine*.

Protocol

NAME
Come importare identificatori OAI-PMH in OpenRefine

CREATED BY
Andrea Marchitelli

PREVIEW

1.1 Aprire OpenRefine e scegliere "Crea progetto"

Software		
OpenRefine		NAME
Freebase, then Google, now open source community		DEVELOPER
https://github.com/OpenRefine/OpenRefine		SOURCE LINK

- 1.2 Selezionare "**Indirizzo/i Web (URLs)**" come fonte dei dati del nuovo progetto e cliccare su "Avanti". Inserire la baseurl OAI-PMH del target individuato, scegliere il verbo **ListIdentifiers** e il metadataPrefix opportuno.
Es. http://www.bibliotecheoggi.it/trends/oai?verb=ListIdentifiers&metadataPrefix=oai_dc

Note

Il protocollo OAI-PMH consente l'impostazione di limitatori per data e per set per esporre dataset parziali e permettere raccolte incrementali.

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

1.3 Selezionare "Files XML" nella sezione "Analizza i dati come"

Screenshot delle opzioni di import in OpenRefine

1.4 Nell'area dell'anteprima del file da importare, individuare la prima riga contenente l'elemento <identifier> e portare il puntatore del mouse a evidenziarla. Poi cliccarvi sopra.

Evidenziare l'identifier

Expected result

A quel punto si sarà ottenuta la lista degli identificatori OAI-PMH, uno per riga, pronta per elaborazioni successive.

Creare le URL per l'harvesting

- 2 Cliccare sulla colonna identifier presente nel progetto e scegliere l'opzione Modifica colonna > Aggiungi colonna basata su questa.

2.1

Nella finestra di dialogo con le impostazioni della nuova colonna inserire **baseurl_oai** come nome della colonna e un'espressione GREL che concateni la baseurl_oai della rivista, il verb GetRecord e l'identificatore OAI già presente nei file.

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

Attenzione a mantenere gli apici (') all'inizio e alla fine della stringa URL

Test del risultato

3

Expected result

A questo punto confermando l'operazione si otterrà una nuova colonna denominata baseurl_oai contenente le URL per il GetRecord di ogni articolo.

97 righe			
Vista a: righe elementi Mostra: 5 10 25 50 righe			
Tutti	Identifier	baseurl_oai	
1.	oai:oji.pkp.sfu.ca:article/37	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/37	
2.	oai:oji.pkp.sfu.ca:article/368	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/368	
3.	oai:oji.pkp.sfu.ca:article/587	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/587	
4.	oai:oji.pkp.sfu.ca:article/718	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/718	
5.	oai:oji.pkp.sfu.ca:article/834	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/834	
6.	oai:oji.pkp.sfu.ca:article/1004	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/1004	
7.	oai:oji.pkp.sfu.ca:article/384	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/384	
8.	oai:oji.pkp.sfu.ca:article/515	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/515	
9.	oai:oji.pkp.sfu.ca:article/627	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/627	
10.	oai:oji.pkp.sfu.ca:article/776	http://www.bibliotecheoggi.it/trends/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oji.pkp.sfu.ca:article/776	

- 4 Testare il risultato ottenuto cliccando su un paio di link e verificando che la pagina corrispondente si apra correttamente