



Nov 30, 2020

© Creare baseurl per harvesting di repository OAI con OpenRefine

Andrea Marchitelli¹

¹Associazione Italiana Biblioteche (AIB)

1 Works for me dx.doi.org/10.17504/protocols.io.bp7pmrmn

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.

EXTERNAL LINK

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

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

EXTERNAL LINK

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

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

Nov 30, 2020

LAST MODIFIED

Nov 30, 2020

PROTOCOL INTEGER ID

45007



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.

BEFORE STARTING

Installare l'applicativo Openrefine

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.



1 1 Aprire OpenRefine e scegliere "Crea progetto"



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

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

1.5



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



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



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