

ISTI Open Access - Feature #20103

Configure ISTI Open Portal to expose usage statistics for OpenAIRE

Nov 12, 2020 09:56 AM - Leonardo Candela

Status:	Completed	Start date:	Feb 09, 2021
Priority:	Normal	Due date:	
Assignee:	Andrea Dell'Amico	% Done:	100%
Category:			
Target version:	ISTI Open Portal Tasks		
Description <p>A follow up of the successful registration of ISTI Open Portal in OpenAIRE (#20053) requires to provide OpenAIRE with repository usage statistics.</p> <p>Please find below the details, I guess we should go for the python script case.</p> <p>OpenAIRE's usage statistic service uses the Matomo Open Source Analytics platform (matomo.org) to track usage activity. When metrics are enabled for a repository, two unique identifiers are generated - a matomo-ID that associates the repository with its usage events in Matomo and an authentication-ID that allows to track usage activity on the Matomo platform. Metadata views and item downloads are tracked and automatically sent to Matomo. Statistics are generated using the COUNTER Code of practice directives.</p> <p>OpenAIRE's usage statistics service tracking code exploits Matomo's API. In order to make the tracking of usage events from repositories more robust, it was necessary to implement repository platform specific patches and plugins starting with DSpace and EPrints. The code is maintained on Github:</p> <ul style="list-style-type: none">• as a patch for various versions of DSpace (https://github.com/openaire/OpenAIRE-Piwik-DSpace)• as an Eprints plugin for version 3 (https://github.com/openaire/EPrints-OAPiwik)• as a python script for all other cases (https://github.com/openaire/Generic-Matomo-Tracker) <p>Going to add the specific configuration parameters into a private area.</p> <p>Details for the configuration files are given in the README of the tracking code.</p>			
Subtasks:			
Feature # 20675: Reconsider the URLs for accessing the payloads			Closed
Related issues:			
Related to ISTI Open Access - Feature #24391: ISPC Open Portal: configure usa...		Closed	Jan 11, 2023
Follows ISTI Open Access - Feature #20053: Register ISTI Open Portal in OpenAIRE		Completed	Nov 02, 2020

History

#2 - Nov 12, 2020 09:57 AM - Leonardo Candela

- Follows Feature #20053: Register ISTI Open Portal in OpenAIRE added

#3 - Dec 28, 2020 05:10 PM - Andrea Dell'Amico

- Status changed from New to In Progress

#4 - Dec 28, 2020 07:58 PM - Andrea Dell'Amico

- % Done changed from 0 to 70

I created the environment and installed the script inside a virtualenv, but it does not work.

It seems that our apache logs directory sports too many files for the script and it gives up after the check of some of the rotated (compressed) files because it does not discriminate between compressed and not compressed files. I opened an issue here <https://github.com/openaire/Generic-Matomo-Tracker/issues/7> asking for a new command line option.

I also see that the script diverged a lot from the original matomo one (that one supports files as arguments).

#5 - Jan 10, 2021 10:48 AM - Leonardo Candela

It seems the revised matomo script has been fixed.

#6 - Jan 10, 2021 07:19 PM - Andrea Dell'Amico

- Status changed from In Progress to Feedback

- % Done changed from 70 to 100

I see. I just installed the script, its execution will be triggered by the logrotate weekly run.

I also just started a run against the one year old access logs.

#7 - Jan 11, 2021 06:19 PM - Andrea Dell'Amico

The import script run without errors, but exactly 0 valid records were found.

The configuration used is the one published in the github repository,

https://github.com/openaire/Generic-Matomo-Tracker/blob/master/matomo_config.yaml, the only modifications are the ID and the auth token.

Can anybody check if the configuration is valid? @michele.artini@isti.cnr.it , @leonardo.candela@isti.cnr.it ?

I guess the relevant part is

```
#OAI-PMH Configuration
oaimh_preamble: "oai:repository.site:"

#Metadata Location (supports multiple locations and regular expressions)
tracking_metadata:
- "/([0-9]*)/(?!pdf)$"

#Item Location (supports multiple locations and regular expressions)
tracking_download:
- "/([0-9]*)/[0-9]*\\w.*pdf"
```

#8 - Jan 11, 2021 08:15 PM - Leonardo Candela

I'm not an expert and, if possible, I'll contact again the OpenAIRE colleagues to get some insights.

I would say that the easiest part is certainly this:

- repository_base_url should be <https://openportal.isti.cnr.it>
- idSite should be 368
- token_auth should be the token

and I'm confident you already set it accordingly.

Regarding the three parameters you identified, it seems they are used by the methods

- check_static
- check_download

to identify the hits in the log to "count". By having a look at the log hits (is it possible to have an excerpt?) we might try to derive them.

For the sake of completeness, oai-pmh records we provide OpenAIRE with are like this

https://openportal.isti.cnr.it/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dnet:people_____:a3e8ec3bad01d7a432cab3cdee473e76

I guess:

- oaimh_preamble might be either
 - the base url of the protocol i.e. <https://openportal.isti.cnr.it/oai>
 - the landing page prefix of the single item <https://openportal.isti.cnr.it/doc>
- tracking_metadata
 - seems the pattern matching expression that, in tandem with the preamble above, identifies an access to the record (by using any url/hit other than ending with ".pdf")
- tracking_download
 - seems the pattern matching expression that, in tandem with the preamble identifies an access to the payload (by using any url/hit ending with ".pdf")

The "problem" that I see is that access to payloads is done via urls like

http://openportal.isti.cnr.it/data/2020/437281/2020_437281_preprint.pdf

i.e. they have a path diverse from that of phm records (/oai) and landing pages (/doc)

An approach might be to consider:

- oaipmh_preamble = <https://openportal.isti.cnr.it>
- tracking_metadata = a regular expression identifying any hit matching the preamble + "/doc" + *
- tracking_download = a regular expression identifying any hit matching the preamble + "/data/" + *

#9 - Jan 12, 2021 01:27 PM - Andrea Dell'Amico

Leonardo Candela wrote:

The "problem" that I see is that access to payloads is done via urls like

http://openportal.isti.cnr.it/data/2020/437281/2020_437281_preprint.pdf

i.e. they have a path diverse from that of phm records (/oai) and landing pages (/doc)

An approach might be to consider:

- oaipmh_preamble = <https://openportal.isti.cnr.it>
- tracking_metadata = a regular expression identifying any hit matching the preamble + "/doc" + *
- tracking_download = a regular expression identifying any hit matching the preamble + "/data/" + *

The oaipmh_preamble should be a piece of a URI and not the hostname. Even an empty string I guess if it's not relevant. About the tracking regular expressions? are we interested in pdf files only?

#10 - Jan 12, 2021 01:55 PM - Andrea Dell'Amico

I tried with:

```
#OAI-PMH Configuration
oaipmh_preamble: "oai?verb"

#Metadata Location (supports multiple locations and regular expressions)
tracking_metadata:
- "/doc?(.*)$"

#Item Location (supports multiple locations and regular expressions)
tracking_download:
- "/data/([0-9]*)/[0-9]*/.*pdf"
```

against a sample of 1000 lines, and something happened:

```
Logs import summary
-----
6 requests imported successfully
1 requests were downloads
5 requests were metadata
701 requests ignored done by bots, search engines...
```

#11 - Jan 12, 2021 01:59 PM - Andrea Dell'Amico

- File *openportal-test.log.gz* added

The log file I used to test the script is attached.

#12 - Jan 12, 2021 03:32 PM - Leonardo Candela

It seems the preamble is not used/useful at all ... none of the log entries contain it :)

Moreover, by looking at the sample of the log I guess this is the entry contributing to the download

93.44.104.144 - - [05/Jan/2020:12:45:03 +0100] "GET /data/2018/395392/2018_395392.postprint.pdf HTTP/1.1" 200 385200 "<https://scholar.google.com/>" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:71.0) Gecko/20100101 Firefox/71.0"

My feeling is that someone discovered the paper via Scholar and accessed it (in fact the paper is indexed by scholar and one of its versions come from openportal).

https://scholar.google.it/scholar?hl=en&as_sdt=0%2C5&q=Do+we+need+new+strategies+for+testing+Systems-of-Systems%3F&btnG=

I would go for the following:

- use an empty string for the preamble;
- the metadata expression is fine;
- the download expression could use any file under the /data/ path ... although the majority of files (if not all) will be pdf;

#13 - Jan 12, 2021 04:46 PM - Andrea Dell'Amico

Leonardo Candela wrote:

It seems the preamble is not used/useful at all ... none of the log entries contain it :)

Not that batch, but there are lots of

```
/oai?verb=....
```

From what I understand they are used to retrieve some metadata

Moreover, by looking at the sample of the log I guess this is the entry contributing to the download

Yes, it's the only valid download

- use an empty string for the preamble;

See above about it.

- the metadata expression is fine;

OK

- the download expression could use any file under the /data/ path ... although the majority of files (if not all) will be pdf;

OK

#14 - Jan 12, 2021 04:54 PM - Andrea Dell'Amico

I just started the import of the year old historic records.

The records of the current week will be imported at the end of the week.

#15 - Jan 12, 2021 05:46 PM - Andrea Dell'Amico

The import of the old data just completed:

Logs import summary

```
71480 requests imported successfully
28856 requests were downloads
42624 requests were metadata
653389 requests ignored done by bots, search engines...
```

Performance summary

```
Total time: 3068 seconds
Requests imported per second: 23.3 requests per second
```

I don't know how to check if they are what we expect they are.

#16 - Jan 12, 2021 06:12 PM - Leonardo Candela

I'm skeptical about the use of the preamble ... it is about the oai-pmh protocol OpenAIRE (and other clients) use to collect all the records metadata ... it is certainly not used to access the files.

The figures we get seems to suggest that, during last year, there were 28856 calls like /oai?verb=.... ... where . is a file name

This sounds strange.

Moreover, access to metadata is done by the oai verb "GetRecord" (thus we should filter the oai calls and not count invocations of verb like ListIdentifier, Identify). Also the ListRecod is problematic because it is an access to the metadata yet in bulk (not for a single record);

Is it possible

- to have a sample of the log where the calls to /oai are recorded ?
- to re-run the script by using an empty preamble ?

#17 - Jan 12, 2021 06:13 PM - Leonardo Candela

Leonardo Candela wrote:

I'm skeptical about the use of the preamble ... it is about the oai-pmh protocol OpenAIRE (and other clients) use to collect all the records metadata ... it is certainly not used to access the files.

The figures we get seems to suggest that, during last year, there were 28856 calls like /oai?verb=.... ... where *.* is a file name

This sounds strange.

Moreover, access to metadata is done by the oai verb "GetRecord" (thus we should filter the oai calls and not count invocations of verb like ListIdentifier, Identify). Also the ListRecod is problematic because it is an access to the metadata yet in bulk (not for a single record);

Is it possible

- to have a sample of the log where the calls to /oai are recorded ?
- to re-run the script by using an empty preamble ?

#18 - Jan 12, 2021 06:22 PM - Andrea Dell'Amico

Leonardo Candela wrote:

I'm skeptical about the use of the preamble ... it is about the oai-pmh protocol OpenAIRE (and other clients) use to collect all the records metadata ... it is certainly not used to access the files.

The figures we get seems to suggest that, during last year, there were 28856 calls like /oai?verb=.... ... where . is a file name

The preamble is not used to build the URL that points to the documents, from what I understand.

This sounds strange.

Moreover, access to metadata is done by the oai verb "GetRecord" (thus we should filter the oai calls and not count invocations of verb like ListIdentifier, Identify). Also the ListRecod is problematic because it is an access to the metadata yet in bulk (not for a single record);

Is it possible

- to have a sample of the log where the calls to /oai are recorded ?

From some weeks ago:

```
194.94.18.130 - - [31/Dec/2020:17:07:53 +0100] "GET /oai HTTP/1.1" 200 6962 "-" "Mozilla/5.0 (Windows NT 6.2; Win64; x64; rv:84.0) Gecko/20100101 Firefox/84.0"
194.94.18.130 - - [31/Dec/2020:17:07:54 +0100] "GET /favicon.ico HTTP/1.1" 200 1585 "https://openportal.isti.cnr.it/oai" "Mozilla/5.0 (Windows NT 6.2; Win64; x64; rv:84.0) Gecko/20100101 Firefox/84.0"
194.94.18.130 - - [31/Dec/2020:17:08:13 +0100] "GET /oai?verb=ListRecords&metadataPrefix=oai_dc HTTP/1.1" 200 60215 "-" "Mozilla/5.0 (Windows NT 6.2; Win64; x64; rv:84.0) Gecko/20100101 Firefox/84.0"
129.70.12.220 - - [31/Dec/2020:17:35:00 +0100] "GET /oai?verb=Identify HTTP/1.1" 200 6762 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:17:35:01 +0100] "GET /oai?verb=ListRecords&metadataPrefix=oai_dc HTTP/1.1" 200 279929 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:17:35:03 +0100] "GET /oai?verb=ListSets HTTP/1.1" 200 6940 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:17:35:04 +0100] "GET /oai?verb=ListMetadataFormats HTTP/1.1" 200 6978 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:32:37 +0100] "GET /oai?verb=Identify HTTP/1.1" 200 6762 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:32:38 +0100] "GET /oai?verb=ListRecords&from=2020-12-31&metadataPrefix=oai_dc HTTP/1.1" 200 6580 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:32:40 +0100] "GET /oai?verb=ListSets HTTP/1.1" 200 6940 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:32:41 +0100] "GET /oai?verb=ListMetadataFormats HTTP/1.1" 200 6978 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:42:20 +0100] "GET /oai?verb=Identify HTTP/1.1" 200 6762 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
```

```

129.70.12.220 - - [31/Dec/2020:18:42:22 +0100] "GET /oai?verb=ListRecords&metadataPrefix=oai_dc HTTP/1.1" 200
279929 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:52.3) Gec
ko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:42:23 +0100] "GET /oai?verb=ListSets HTTP/1.1" 200 6940 "-" "Mozilla/5.0 (X1
1; Linux x86_64; rv:52.3) Gecko/20100101 Firefox/52.3"
129.70.12.220 - - [31/Dec/2020:18:42:25 +0100] "GET /oai?verb=ListMetadataFormats HTTP/1.1" 200 6978 "-" "Mozi
lla/5.0 (X11; Linux x86_64; rv:52.3) Gecko/20100101 Fire
fox/52.3"
213.135.60.121 - - [01/Jan/2021:08:18:05 +0100] "GET /oai?verb=ListRecords&metadataPrefix=oai_dc&from=2020-12-
23 HTTP/1.1" 200 279753 "-" "Mozilla/5.0 (compatible; OA
I; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:11 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C100%7C5f9a924d2970f022109
6d967%7Ctrue%7C HTTP/1.1" 200 281410 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:12 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C200%7C5f9a924e2970f022109
6d9cb%7Ctrue%7C HTTP/1.1" 200 277039 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:12 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C300%7C5f9a924e2970f022109
6da2f%7Ctrue%7C HTTP/1.1" 200 290857 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:13 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C400%7C5f9a924e2970f022109
6da93%7Ctrue%7C HTTP/1.1" 200 271118 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:13 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C500%7C5f9a924e2970f022109
6daf7%7Ctrue%7C HTTP/1.1" 200 297614 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:14 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C600%7C5f9a924f2970f022109
6db5b%7Ctrue%7C HTTP/1.1" 200 303997 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:14 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C700%7C5f9a924f2970f022109
6dbbf%7Ctrue%7C HTTP/1.1" 200 263452 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:15 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C800%7C5f9a924f2970f022109
6dc23%7Ctrue%7C HTTP/1.1" 200 233275 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"
213.135.60.121 - - [01/Jan/2021:08:18:15 +0100] "GET /oai?verb=ListRecords&resumptionToken=-1%7Coai_dc%7C*+AND
+datestamp+%3E%3D+2020-12-23%7C900%7C5f9a92502970f022109
6dc87%7Ctrue%7C HTTP/1.1" 200 265360 "-" "Mozilla/5.0 (compatible; OAI; +http://www.openaire.eu)"

```

- to re-run the script by using an empty preamble ?

This can lead to duplicate records in matomo.

#19 - Jan 12, 2021 08:31 PM - Leonardo Candela

Andrea Dell'Amico wrote:

The import of the old data just completed:

Logs import summary

```

71480 requests imported successfully
28856 requests were downloads
42624 requests were metadata
653389 requests ignored done by bots, search engines...

```

Performance summary

```

Total time: 3068 seconds
Requests imported per second: 23.3 requests per second

```

I don't know how to check if they are what we expect they are.

According to Google Analytics, in the period **01 Jan 2020 - 31 Dec. 2020** the open portal website had

- 27,774 pageviews
- 21,742 unique pageviews

This page <https://openportal.isti.cnr.it/doc?id=people::b3e08179b448c1d31dfe491e006da982> get

- 2,291 pageviews

- 1,917 unique page views

#20 - Jan 12, 2021 10:36 PM - Andrea Dell'Amico

Leonardo Candela wrote:

Andrea Dell'Amico wrote:

The import of the old data just completed:

Logs import summary

```
-----  
71480 requests imported successfully  
28856 requests were downloads  
42624 requests were metadata  
653389 requests ignored done by bots, search engines...
```

Performance summary

```
-----  
Total time: 3068 seconds  
Requests imported per second: 23.3 requests per second
```

I don't know how to check if they are what we expect they are.

According to Google Analytics, in the period **01 Jan 2020 - 31 Dec. 2020** the open portal website had

I was referring to the openaire usage statistics site :-). I don't think I have access, and even than I don't know how to check if the collected information is correct.

#21 - Jan 12, 2021 11:39 PM - Leonardo Candela

According to my understanding the figures we are going to store in matomo are usage statistics that are independent from OpenAIRE, i.e. they are usage statistics measured and declared by the single repository.

OpenAIRE is going to complement them with its own figures.

I was reporting the Google Analytics figures to highlight that the repository usage statistics we are going to store in matomo seems not comparable with usage statistics GA collect about our repository.

Was the script analysing only 2020 logs? Could you please let me know the time frame of the log we are analysing by the matomo script?

#22 - Jan 13, 2021 02:02 PM - Andrea Dell'Amico

Leonardo Candela wrote:

According to my understanding the figures we are going to store in matomo are usage statistics that are independent from OpenAIRE, i.e. they are usage statistics measured and declared by the single repository.

OpenAIRE is going to complement them with its own figures.

I try to rephrase: *Do we have access to the OpenAIRE matomo data, to see what was imported?*

I was reporting the Google Analytics figures to highlight that the repository usage statistics we are going to store in matomo seems not comparable with usage statistics GA collect about our repository.

I don't think they are meant to be comparable?

Was the script analysing only 2020 logs? Could you please let me know the time frame of the log we are analysing by the matomo script?

Oldest entry:

```
66.249.76.123 - - [05/Jan/2020:06:25:27 +0100] "GET /results?format=xml&pageSize=16&cnrcreatorforbrowsing%5B%5D=Antonia%20Bertolino&rights%5B%5D=Open%20Access HTTP/1.1" 200 23067 "-" "Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
```

Newest one:

```
185.191.171.45 - - [11/Jan/2021:06:25:08 +0100] "GET /results?id=people_____%3A%3A0f040e21ccdf16af3624f3719159dalf&option=com_dnetindexclient&view=doc HTTP/1.1" 200 10194 "-" "Mozilla/5.0 (compatible; SemrushBot/7~bl; +http://www.semrush.com/bot.html) "
```

#23 - Jan 13, 2021 06:03 PM - Leonardo Candela

Andrea Dell'Amico wrote:

Leonardo Candela wrote:

According to my understanding the figures we are going to store in matomo are usage statistics that are independent from OpenAIRE, i.e. they are usage statistics measured and declared by the single repository.

OpenAIRE is going to complement them with its own figures.

I try to rephrase: *Do we have access to the OpenAIRE matomo data, to see what was imported?*

I don't think so. Moreover, I decided to send an email to OpenAIRE colleagues (you are in cc).

I was reporting the Google Analytics figures to highlight that the repository usage statistics we are going to store in matomo seems not comparable with usage statistics GA collect about our repository.

I don't think they are meant to be comparable?

Why? I would expect that the figures we get from GA about page views are greater than the figures we publish for "metadata view" ... at the end of the story a "metadata view" is a view of a specific web page.

#24 - Jan 29, 2021 08:13 AM - Leonardo Candela

- % Done changed from 100 to 90

According to the information we received from OpenAIRE colleagues we should reconsider our config as follows

oaipmh_preamble: "oai:dnet:people____::"

#25 - Feb 17, 2021 06:47 PM - Andrea Dell'Amico

I reconfigured the tool and manually run the processing of the last week logs. The result:

Logs import summary

```
899 requests imported successfully
610 requests were downloads
289 requests were metadata
9323 requests ignored done by bots, search engines...
```

Performance summary

```
Total time: 13 seconds
Requests imported per second: 68.82 requests per second
```

#26 - Apr 06, 2021 10:50 AM - Leonardo Candela

It seems that the current config is not working, this is the message we get from OpenAIRE team:

I 've looked at the usage activity from ISTI and I 've noticed that we are receiving two different patterns for ISTI records:

1. For downloads: oai:dnet:people____::2018
2. For pageviews: oai:dnet:people____::?id=people____::33ea8dd21185e71591a6ca43951ea24c

We should alter the Matomo config I guess. The pageviews is correct ... the downloads part should count URLs like this

http://openportal.isti.cnr.it/data/2021/428210/2020_%20428210.preprint.pdf?id=people_____%3A%3A7362a3828ba52cbea14949c31be9def0

#27 - Apr 06, 2021 11:33 AM - Andrea Dell'Amico

Our matomo tracker configuration reads:


```
#OAI-PMH Configuration
oaipmh_preamble: "oai:dnet:people_____: "

#Metadata Location (supports multiple locations and regular expressions)
tracking_metadata:
- "/doc?(.*)$"

#Item Location (supports multiple locations and regular expressions)
tracking_download:
- "/data/([0-9]*)/[0-9]*/.*"

```

#28 - Apr 08, 2021 01:27 PM - Leonardo Candela

- *File matomo_import_logs.py added*
- *File matomo_config.yaml added*

We have received a revised version of the scripts (in attachment) ... I kindly ask @andrea.dellamico@isti.cnr.it to install them.

#29 - Apr 08, 2021 06:03 PM - Andrea Dell'Amico

Done.

#30 - Apr 16, 2021 07:15 PM - Leonardo Candela

- *Status changed from Feedback to Completed*

We received a positive feedback from OpenAIRE colleagues on the correctness of the information we provide them with.

#31 - Jan 11, 2023 11:30 AM - Leonardo Candela

- *Related to Feature #24391: ISPC Open Portal: configure usage counts added*

Files

openportal-test.log.gz	26.7 KB	Jan 12, 2021	Andrea Dell'Amico
matomo_config.yaml	1.51 KB	Apr 08, 2021	Leonardo Candela
matomo_import_logs.py	56.1 KB	Apr 08, 2021	Leonardo Candela