Towards a Network of Usage Statistics Hubs on a Global Level

Alternative Indicators Summer School
September 2018

Andreas Czerniak, Bielefeld University
Dimitris Pierrakos, ATHENA Research Center
Jochen Schirrwagen, Bielefeld University
TOPICS

OpenAIRE infrastructure and Usage Statistics Service
  Usage Stats sketch, How it works, Tracking Workflow, Service Features.

Metrics in the OpenAIRE Content Provider Dashboard
  When and How to participate

OpenAIRE: a usage statistics Hub for Responsible Metrics
  Main Challenges and Limitations, Next steps
OpenAIRE 2020

- A pan-European Research Information platform to monitor OA research outcomes from EC and other national funders.
- Research analytics tools to promote new scientific metrics & support evidence-based decision-making.
- Implementation of an OpenAIRE usage statistics service for usage data collected from data providers.
OpenAIRE-Advance

• Continues the mission of OpenAIRE aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud.

• Empowers National Open Access Desks (NOADs) so they become a pivotal part within their own national data infrastructures, positioning OA and open science onto national agendas.

• Improves the OpenAIRE services:
  • optimizing their performance and scalability.
  • refining their functionality based on end-user feedback.
  • repackaging them into products, taking a professional marketing approach with well-defined KPIs.
  • consolidating the range of services/products into a common e-Infra catalogue to enable a wider uptake.

• Consolidates OpenAIRE’s global role extending its collaborations with Latin America, US, Japan, Canada, and Africa.
Usage Statistics in OpenAIRE-Advance

**Task covers:**
- aligning policies and standards for gathering and sharing of usage data -> guidelines
- considering legal aspects (GDPR compliance)
- relating usage statistics to other kinds of metrics
- collecting and processing of usage data and producing consolidated, standards-based usage statistics

**Task team:**
Athena Research Center, University of Bielefeld, University of Minho, Jisc IRUS-UK, Couperin + NOADs
OpenAIRE infrastructure and Usage Statistics Service

Usage Stats sketch, How it works, Tracking Workflow, Service Features.
Usage Statistics Service Features

Tracking of views and downloads / collecting COUNTER reports
Push or Pull collection workflows.

Anonymisation of IP-addresses.

Metadata de-duplication enables accumulation of views and downloads for same documents

COUNTER Code of Practice compatibility.
standards based usage statistics.
enables comparability with statistics from other data sources.
Matomo Analytics platform

- World's leading open-source analytics platform (formerly Piwik).
- Valuable insights into website traffic and visitors activity.
- Matomo collects and stores PII (personally identifiable information).
- Keeps full data ownership and can control who has access.
- Robot filtering plugin.
- Compliant with EU regulations.
- Recommended by privacy organizations such as ULD (Germany) and CNIL (France).
## Matomo Facts

<table>
<thead>
<tr>
<th></th>
<th>Matomo</th>
<th>Google Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hits per Month</td>
<td>Unlimited</td>
<td>10 million</td>
</tr>
<tr>
<td>Number of user accounts per login</td>
<td>Unlimited</td>
<td>10</td>
</tr>
<tr>
<td>Data storage time</td>
<td>Unlimited</td>
<td>25 months</td>
</tr>
<tr>
<td>Number of properties (websites, apps etc.) tracked per account</td>
<td>Unlimited</td>
<td>50</td>
</tr>
<tr>
<td>Custom Variables</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Data Export</td>
<td>Unlimited</td>
<td>5000 rows</td>
</tr>
<tr>
<td>Real time Analytics</td>
<td>Matomo offers real-time web analytics in all of its reports.</td>
<td>GA monitors user activity right after it happens, although period of delay is not explicitly stated.</td>
</tr>
</tbody>
</table>
## Matomo Tracking Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>idSite</td>
<td>the ID of the repository</td>
</tr>
<tr>
<td>idVisit</td>
<td>a visitor/session ID (an 8 byte binary string)</td>
</tr>
<tr>
<td>visitIP (optionally anonymized)</td>
<td>the IP address of the visitor</td>
</tr>
<tr>
<td>action</td>
<td>the action performed (view, download, outlink, etc)</td>
</tr>
<tr>
<td>url</td>
<td>the url of the requested item</td>
</tr>
<tr>
<td>timestamp</td>
<td>the date &amp; time of the request</td>
</tr>
<tr>
<td>OAI-PMH Identifier</td>
<td>the Open Access Initiative identifier of the item being viewed/downloaded</td>
</tr>
<tr>
<td>agent</td>
<td>the Web Browser and the operating system of the visitor</td>
</tr>
<tr>
<td>referrer</td>
<td>The url linked to the item requested</td>
</tr>
</tbody>
</table>
Tiers Collection Workflows for Usage Statistics

PUSH
tracked event

Matomo
processing script

Metadata-Index

UsageStatistics-DB

PULL
COUNTER Report

processing script

Repository
- CRIS
- eJournal

National Statistics Node
- Publisher

IP-Anonym.
An institutional repository is registered in Matomo.

Server side tracking: Plugins (Dspace) or patches (Eprints) using Matomo’s HTTP API.

Usage Activity is tracked and logged at Matomo platform in real time.

Information is transferred offline, using Matomo’s API, to OpenAIRE’s DBs for statistical analysis.

Statistics are deployed for human (OpenAIRE’s Portal) and machine (Sushi-Lite API) consumption.
Collecting (Pull) Consolidated Usage Statistics Reports

- Gathering of **consolidated statistics reports from aggregation services**, such as IRUS-UK, using protocols such as SUSHI-Lite.
- Statistics are **stored** to OpenAIRE’s DB for statistical analysis.
- Statistics are **deployed** via OpenAIRE’s Portal or Sushi-Lite API.
Usage Statistics in the OpenAIRE Infrastructure

- OpenAIRE collects from > 1000 compatible data providers
  ~24 Mio documents
- currently 99 active data providers participating in Usage statistics + IRUS-UK
- Usage statistics deployment under CC-0 license.
  - in OpenAIRE dashboard, portal and API.
Usage Activity in Real Time

Monday, November 28, 2016 - 23:56:01
IP: 177.201.227.142
Provider: Unknown

Google

2 Actions - 2s
1. Biblioteca Digital do IPB: As paisagens do desenho: características formais do desenho de paisagens
   https://bibliotecadigital.ipb.pt/handle/10198/6333

Monday, November 28, 2016 - 23:54:15
IP: 191.7.209.66
Provider: Online

Website: scholar.google.com.br

3 Actions - 1 min 46s
1. Biblioteca Digital do IPB: Utilização do software de geometria dinâmica GeoGebra por alunos do 3º Ciclo do Ensino...
   https://bibliotecadigital.ipb.pt/handle/10198/10401
Real Time Visitor Map
Usage Activity Statistics

Number of Downloads

- IRUS-UK
- OpenAIRE(Metamo)
- SARC-OJS

Yearly Breakdown:
- 2016
- 2017
- 2018
Usage Activity Statistics

Number of Views (only by Matomo)

- 2016: 2000000
- 2017: 4000000
- 2018: 1500000
Usage Activity Statistics
Usage Activity Statistics

Downloads & Views per Item type

- Article: Downloads > Views
- Doctoral thesis: Downloads, Views
- Dissertation: Downloads, Views
- Book: Downloads, Views
- Conference object: Downloads, Views
- Part of book or chapter of book: Downloads, Views
- Master thesis: Downloads, Views
- QoC: Downloads, Views
- Research: Downloads, Views
- Report: Downloads, Views
- Biological thesis: Downloads, Views
- Preprint: Downloads, Views
- External research report: Downloads, Views
- Lecture: Downloads, Views
- Review: Downloads, Views
- Patent: Downloads, Views
- Collection: Downloads, Views
- Annotation: Downloads, Views
- Audiovisual: Downloads, Views
- Image: Downloads, Views
- Software: Downloads, Views
2

Metrics in the OpenAIRE Content Provider Dashboard
When and how to participate?
one stop shop for OpenAIRE content providers

for friends...
“the repository managers dashboard”

Dashboard for content providers
Register
Register data sources in the OpenAIRE infrastructure

Validate
Validate data sources against OpenAIRE guidelines

Notifications
View notifications to enrich the metadata and the content

Metrics
View aggregated, cleaned usage statistics for repository access
Metrics in the OpenAIRE Dashboard for content providers

- Four steps to join OpenAIRE Usage Statistics
  1. Download.
  2. Configure.
  3. Deploy.
  4. Validate (by OpenAIRE).

- Or enter SUSHI endpoint to let OpenAIRE collect COUNTER reports
Enable Metrics for selected Datasource

Metrics

YOU DON'T HAVE METRICS ENABLED FOR THIS REPOSITORY YET. WOULD YOU LIKE TO ENABLE THEM?

Once you select to enable metrics for your repository, the following steps need to be performed:

On your side
1. Download the tracking code for your repository platform
2. Configure the tracking code according to the instructions
3. Deploy the tracking code in your repository platform

On the OpenAIRE's side
4. Validate the installation of the tracking code and inform the repository manager accordingly

For more details about the workflows and tools please consult the "Guidelines for Collecting Usage Events and Provision of Usage Statistics".
## Configure Metrics for selected Datasource

OpenAIRE's usage statistic service uses the Piwik Open Source Analytics platform (piwik.org) to track usage activity. When metrics are enabled for a repository, two unique identifiers are generated - a piwik-ID that associates the repository with its usage events in Piwik and an authentication-ID that allows to track usage activity on the Piwik platform. Metadata views and item downloads are tracked and automatically sent to Piwik. Statistics are generated using the COUNTER Code of practice directives.

OpenAIRE’s usage statistics service tracking code exploits Piwik’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:

- as a patch for various versions of DSpace ([https://github.com/openaire/OpenAIRE-Piwik-DSpace](https://github.com/openaire/OpenAIRE-Piwik-DSpace))

To configure your repository to allow tracking in Piwik platform, please change the configuration files with the following parameters and values, generated for your site:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PiwikID</td>
<td>000</td>
</tr>
<tr>
<td>AuthenticationToken</td>
<td>01233456</td>
</tr>
</tbody>
</table>

Details for the configuration files are given in the README of the tracking code.
Usage Metrics by example of the UMinho Repository
In this paper, we show, from the consumer's budget model, that consumption, aggregate wealth, and labour income yields. We use data for several OECD countries and find that they temporarily lower consumption to rise. The component of asset wealth, then investors react in the first quarter, signalling a future rise in taxes, then they will temporarily lower consumption to rise.
SUSHI-Lite Interface

- Available as beta with the help of IRUS-UK
  - [http://beta.services.openaire.eu/usagestats/sushilite/](http://beta.services.openaire.eu/usagestats/sushilite/)

- Supports COUNTER R4 compatible reports:
  - Article Reports (AR) and Book Reports (BR) using identifiers like openaire, doi, oai-record-id
  - Item Reports (IR)
  - Repository Reports (RR) using identifiers issued by OpenAIRE or OpenDOAR
  - Journal Reports (JR) using identifiers like ISSN

COUNTER USAGE REPORTS
RELEVANT FOR OPENAIRE

- Repository Platform Report – RR1
- Journal Report – JR1
- Item Report – IR1
- Article Report – AR1
- Book Report – BR1
- Book Chapter Report – BR2
- Research Dataset Report tbc.
SUSHI response example (JSON)

Repository Report

```
- {  
  - {    
    ItemIdentifier: {}  
  }  
  - {    
    Type: "OpenAIRE",    
    Value: "openAIRE:8e98d81f8217394975cccb3337bb5761"  
  }  
  - {    
    Type: "OpenDOAR",    
    Value: "307"  
  }  
  - {    
    Type: "URL",    
    Value: "https://repositorium.sdum.uminho.pt/"  
  }  
},  
ItemPlatform: "Universidade do Minho: RepositoriUM",  
ItemDataType: "Platform",  
ItemPerformance: [  
  - {    
    Period: {      
      Begin: "2017-01-01",      
      End: "2017-01-31"    
    },    
    Instance: [      
      {        
        MetricType: "ft_total",        
        Count: "22087"      
      },      
      {        
        MetricType: "abstract",        
        Count: "51685"      
      }    
    ]  
  },  
  Category: "Requests"  
},
```

Item Report

```
Report: {  
  @Created: "2017-09-06 08:08:21+0000",  
  @Version: "4",  
  @Name: "IRI-4",  
  - Vendor: {    
    Contact: {      
      Name: "OpenAIRE"    
    },    
    E-mail: "helpdesk@openaire.eu"  
  },  
  - Customer: {    
    ID: "anonymous",    
    - ReportItems: [      
        {          
          - ItemIdentifier: {            
            Type: "openAIRE",            
            Value: "dodug_v7.001:8332b24d03f7f07285c08895650189489"          
          },          
          - ItemPlatform: "IRI-4",          
          Value: "http://hdl.handle.net/1832/7975; http://hdl.handle.net/1832/7463; http://europepmc.org/articles/PMC2208319;"          
          },          
          - ItemDataType: "Article",          
          - ItemData: "Adaptive evolution of a lactose-consuming Saccharomyces cerevisiae recombinant",          
          - ItemPerformance: [            
              {                
                Period: {                  
                  Begin: "2017-01-01",                  
                  End: "2017-01-31"                
                },                
                Instance: [                  
                    {                      
                      MetricType: "ft_total",                      
                      Count: "5"                  
                    },                  
                    {                      
                      MetricType: "abstract",                      
                      Count: "4"                  
                    }                
                ]            
          }          
      ]    
  },  
  Category: "Requests"  
},
```
OpenAIRE: a usage statistics Hub for Responsible Metrics

Main Challenges and Limitations, Next steps
Main Challenges and Limitations

- **Comparability of Usage Statistics across platforms**
  - Must be collected and processed by agreed common standards, e.g. COUNTER CoP
  - Must be openly accessible by default (e.g. CC-0 license)

- **Comparability of Usage Statistics of different versions of an item**
  - E.g. Usage Statistics of Open Access vs. non-Open Access item versions
  - Usage Statistics must be enriched by item identifiers
  - Items must have comprehensive metadata descriptions

- **Contextualized Usage Statistics**
  - COUNTER Reports are a basic first step but limited on statistics per platform (items)
  - Aggregated usage statistics of deduplicated items
  - Linked research results and their Usage Statistics in the context of (e.g. a project or topic)

"The larger the network the better the usage statistics"
Next Steps

- Develop Matomo plugins for other Repository platforms (eg. Fedora, Samvera)
- Promote the service to content provider managers
- Support national usage statistics initiatives to become a node in OpenAIRE Usage Statistics
  - E.g. LA Referencia to set up a regional usage statistics network
- Contribute to the Open Metrics concept and vision
- Update to COUNTER R5 standard and support of COUNTER CoP for Research Data
The Usage Statistics Hub Concept

- We already link publications and datasets
- Let’s link their usage statistics too, make usage statistics more valuable and meaningful
www.openaire.eu
@openaire_eu
facebook.com/groups/openaire
andreas.czerniak@uni-bielefeld.de
dpierrakos@gmail.com
jochen.schirrwagen@uni-bielefeld.de