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

OpenAIRE

European Commission

@openaire_eu

TOPICS
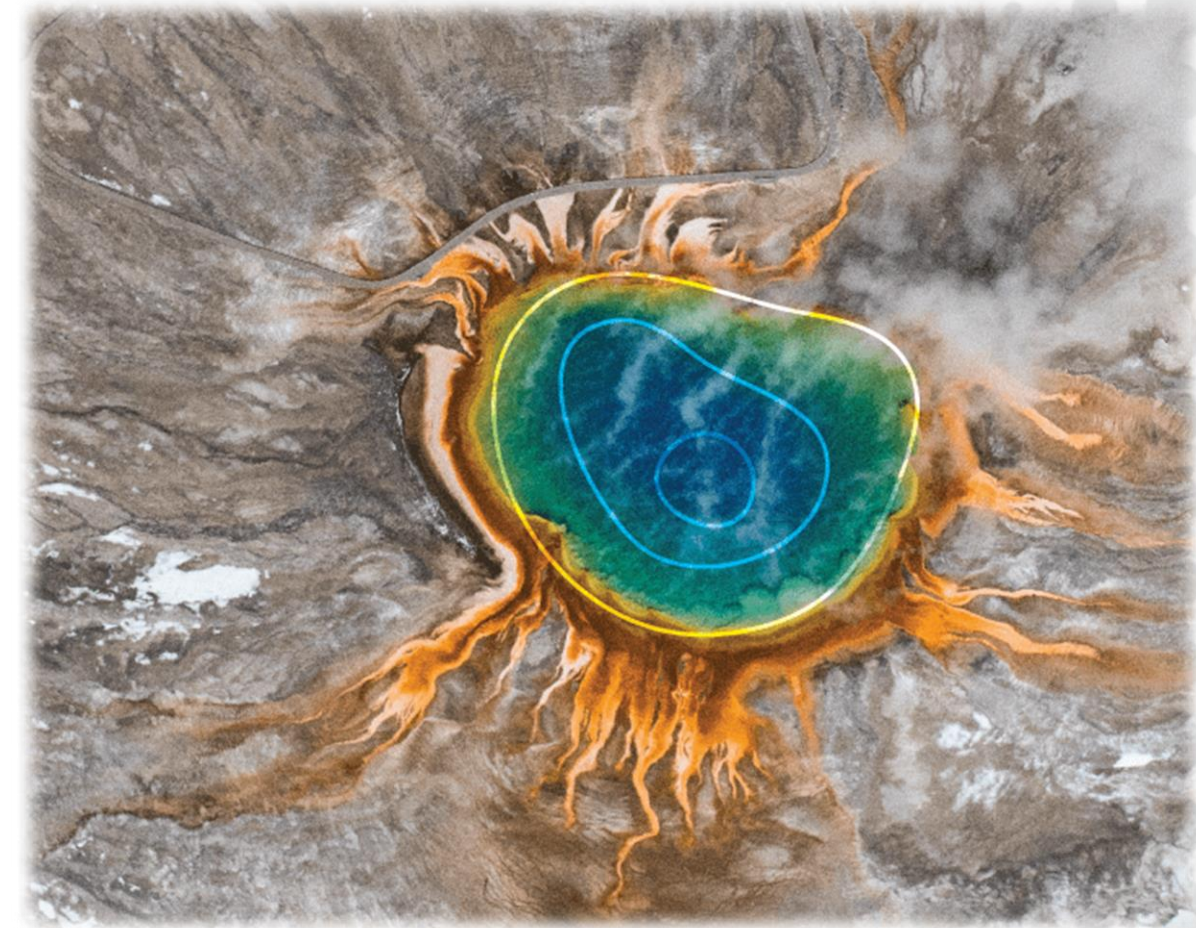
OpenAIRE infrastructure and Usage Statistics Service

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

www.openaire.eu

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   European Commission

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.

OpenAIRE

European Commission

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

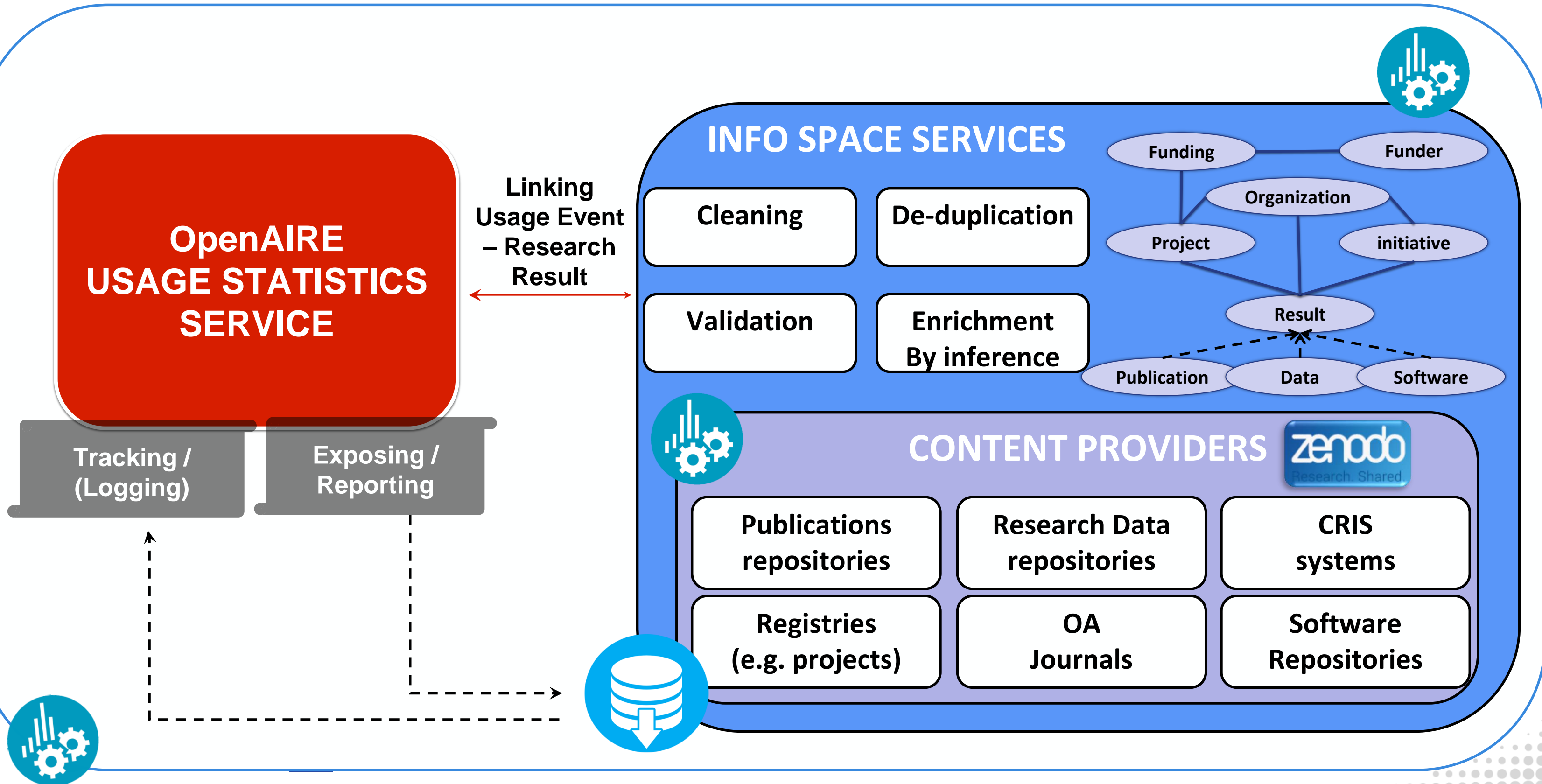OpenAIRE   European Commission

1

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

OpenAIRE's e-infrastructure Commons & Usage Stats sketch

OpenAIRE USAGE STATISTICS SERVICE

Tracking / (Logging)

Exposing / Reporting

Linking Usage Event – Research Result

INFO SPACE SERVICES

Funding

Funder

Organization

Project

initiative

Result

Cleaning

De-duplication

Validation

Enrichment By inference

Publication

Data

Software

CONTENT PROVIDERS

zenodo
Research. Shared.

Publications repositories

Research Data repositories

CRIS systems

Registries (e.g. projects)

OA Journals

Software Repositories

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

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

Matomo Tracking Parameters

Parameter	Description
idSite	the ID of the repository
idVisit	a visitor/session ID (an 8 byte binary string)
visitIP (optionally anonymized)	the IP address of the visitor
action	the action performed (view, download, outlink, etc)
url	the url of the requested item
timestamp	the date & time of the request
OAI-PMH Identifier	the Open Access Initiative identifier of the item being viewed/downloaded
agent	the Web Browser and the operating system of the visitor
referrer	The url linked to the item requested

Tiers Collection Workflows for Usage Statistics
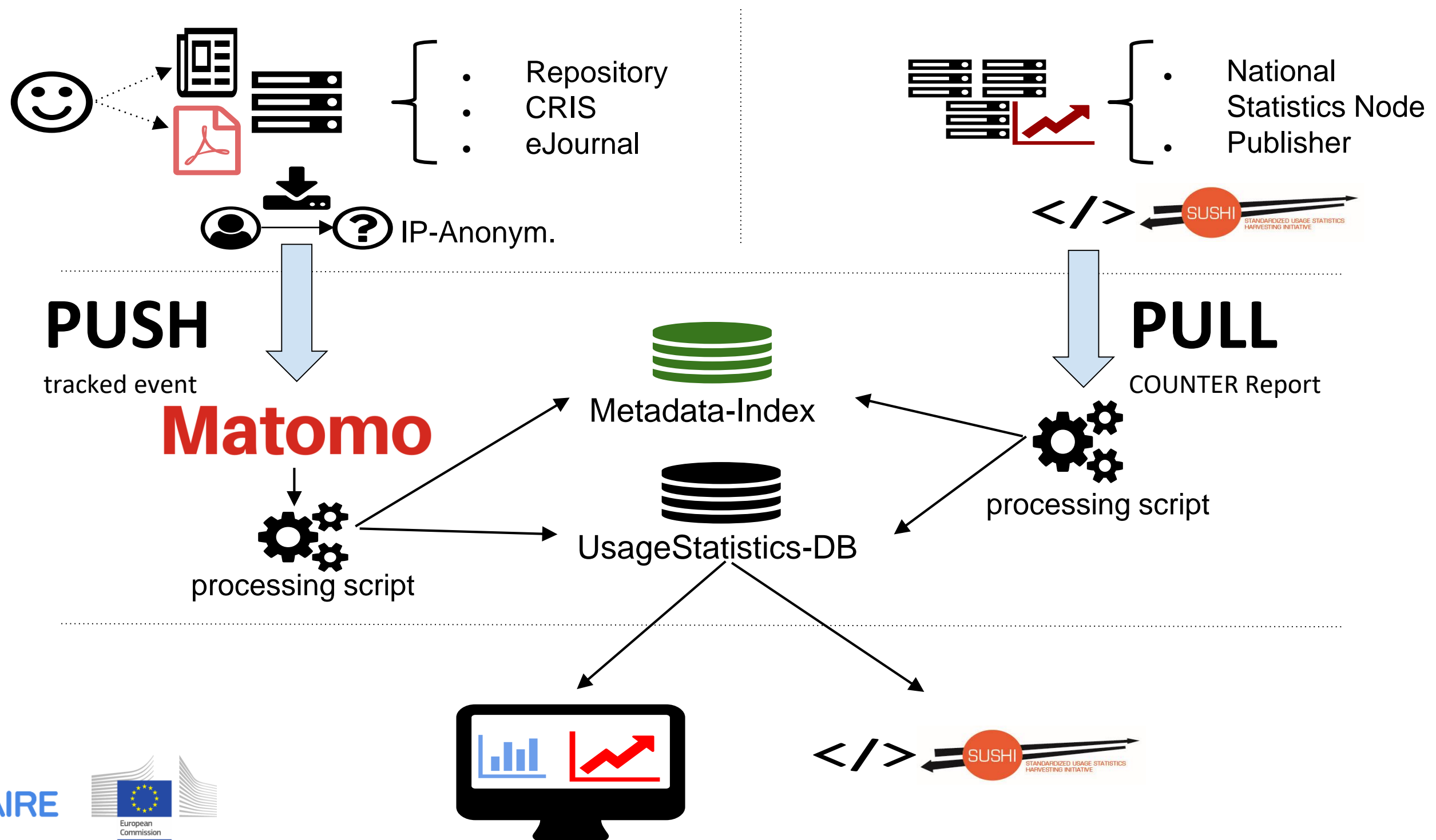
Push Usage Statistics Tracking Workflow

TIER-1

- 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.

OpenAIRE

European Commission
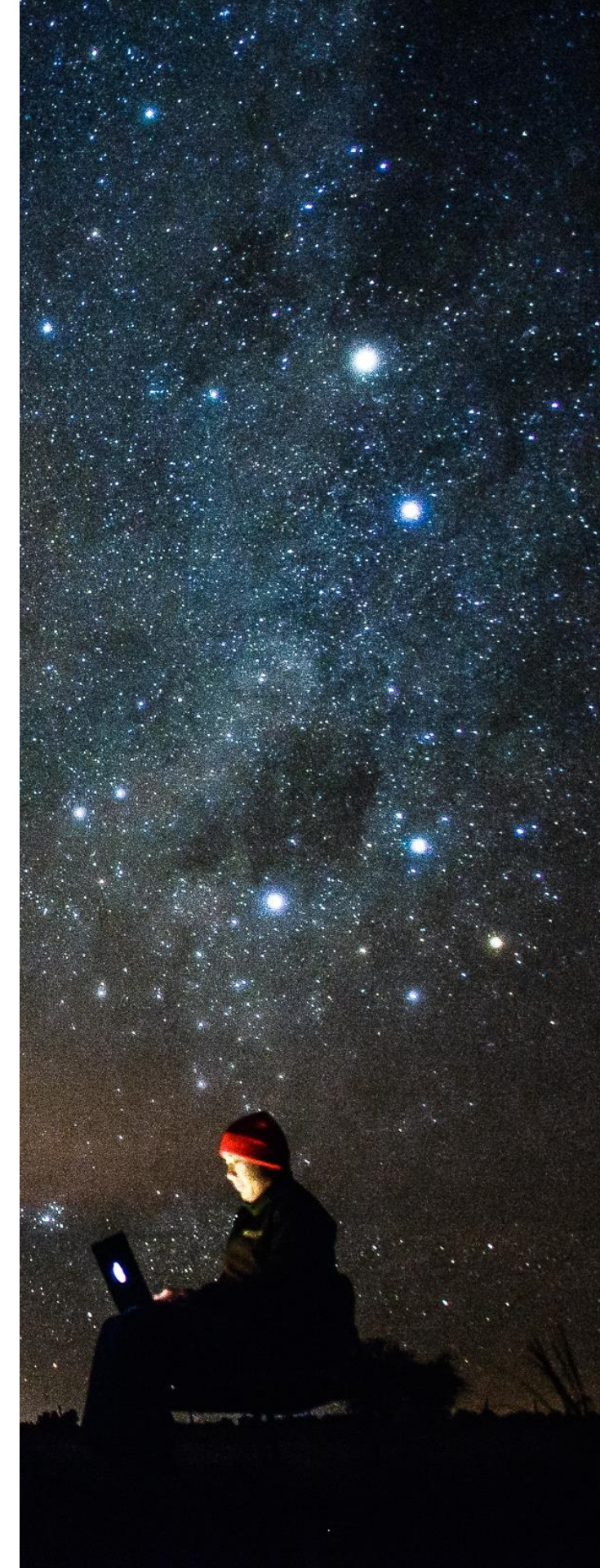
Collecting (Pull) Consolidated Usage Statistics Reports

TIER-2

- 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.

OpenAIRE
European Commission
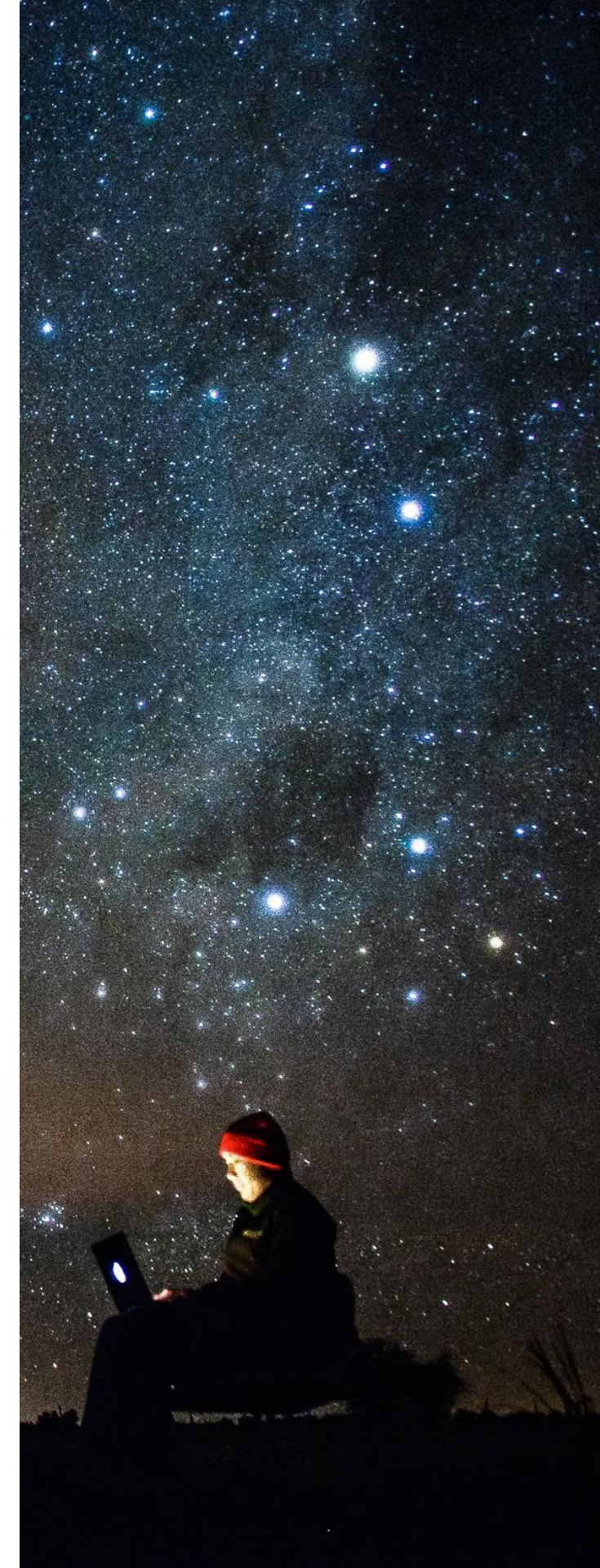
OpenAIRE Usage Statistics DB
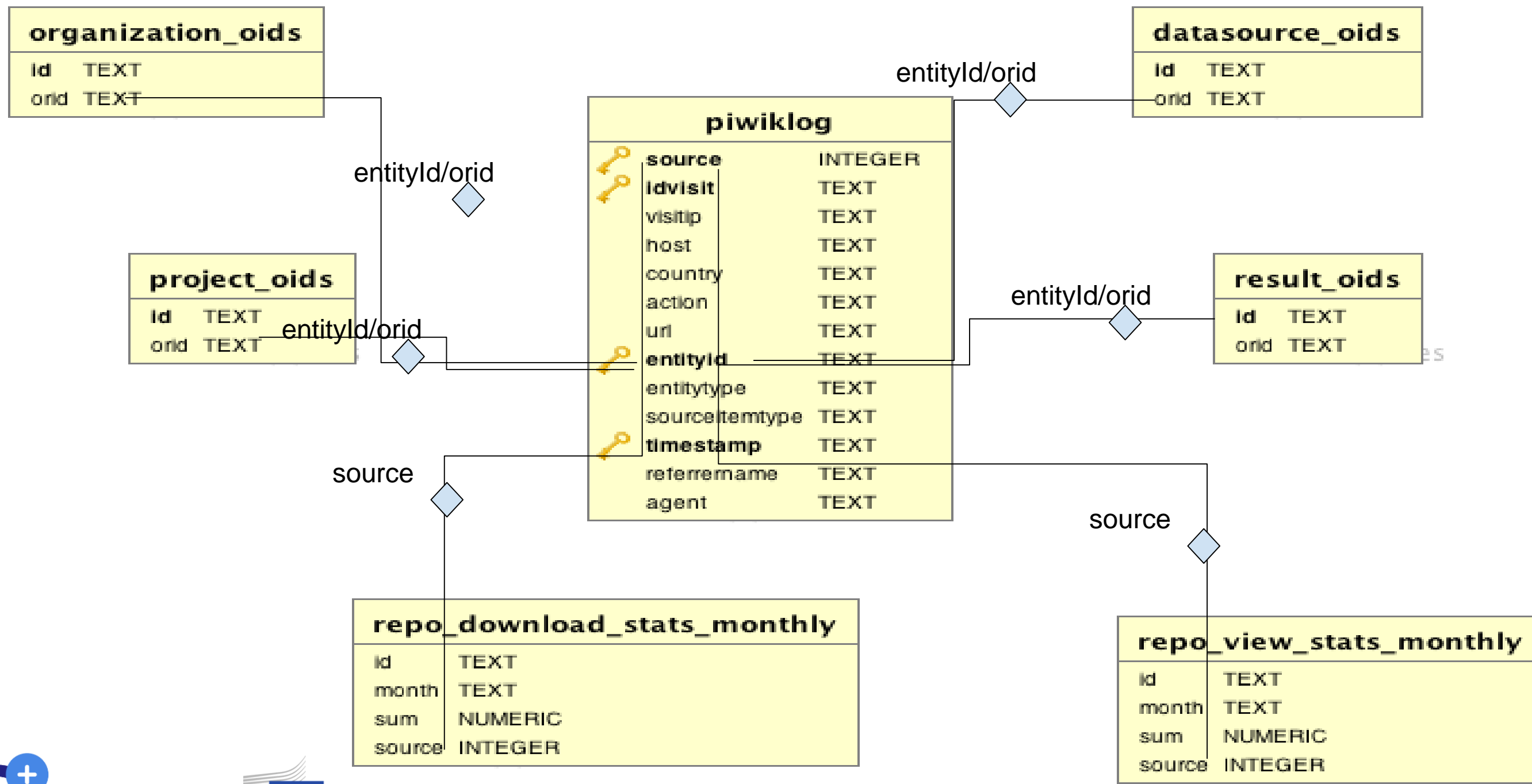
Usage Statistics in the OpenAIRE Infrastructure

- **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. **2x** 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: 191.7.209.86
Provider: Online
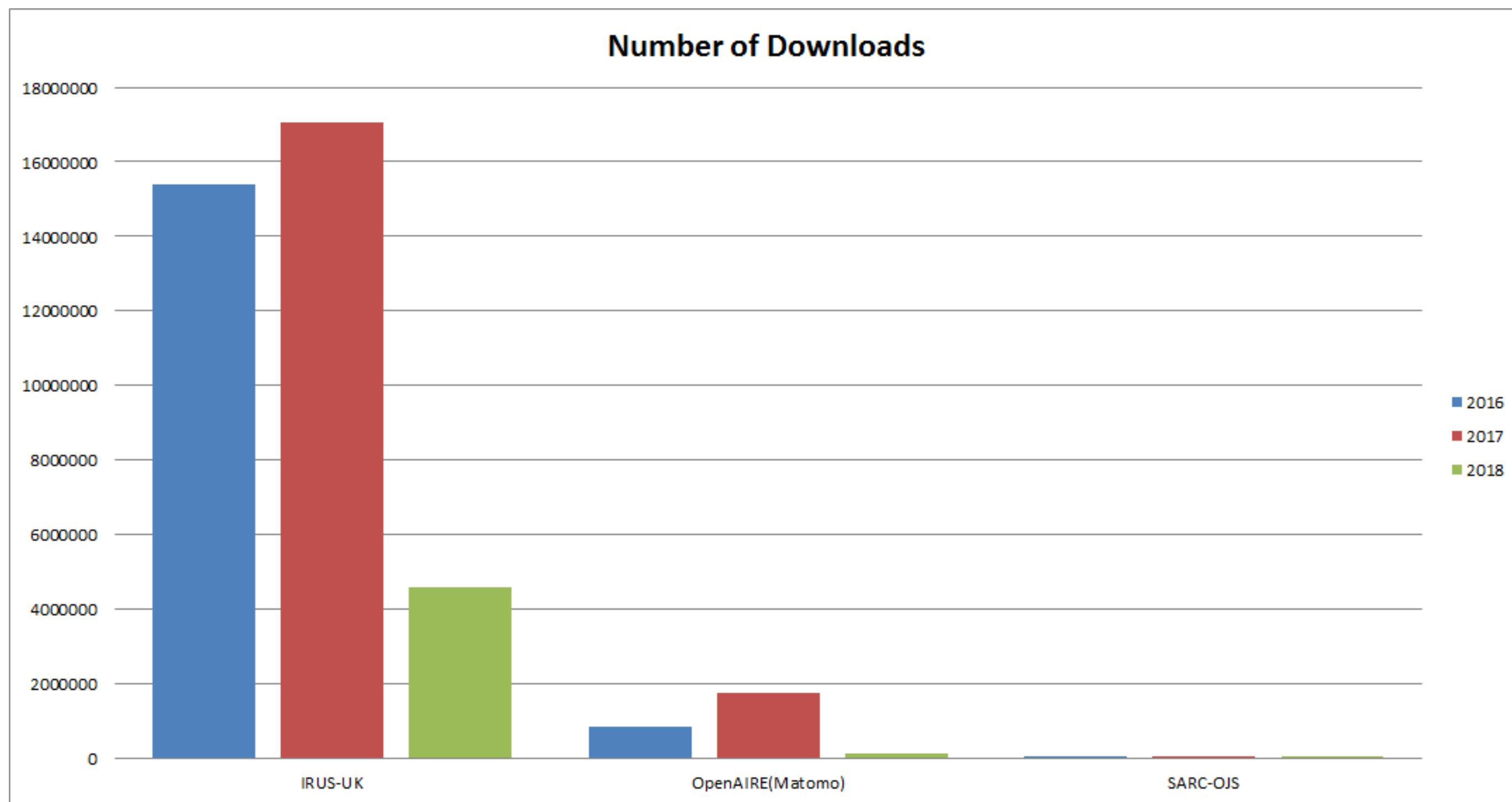
Website: scholar.google.com.br

3 Actions - 1 min 46s
1. **2x** 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
3. https://bibliotecadigital.ipb.pt/bitstream/10198/10401/1/PA_2013ieticSalamanca.pdf

Real Time Visitor Map
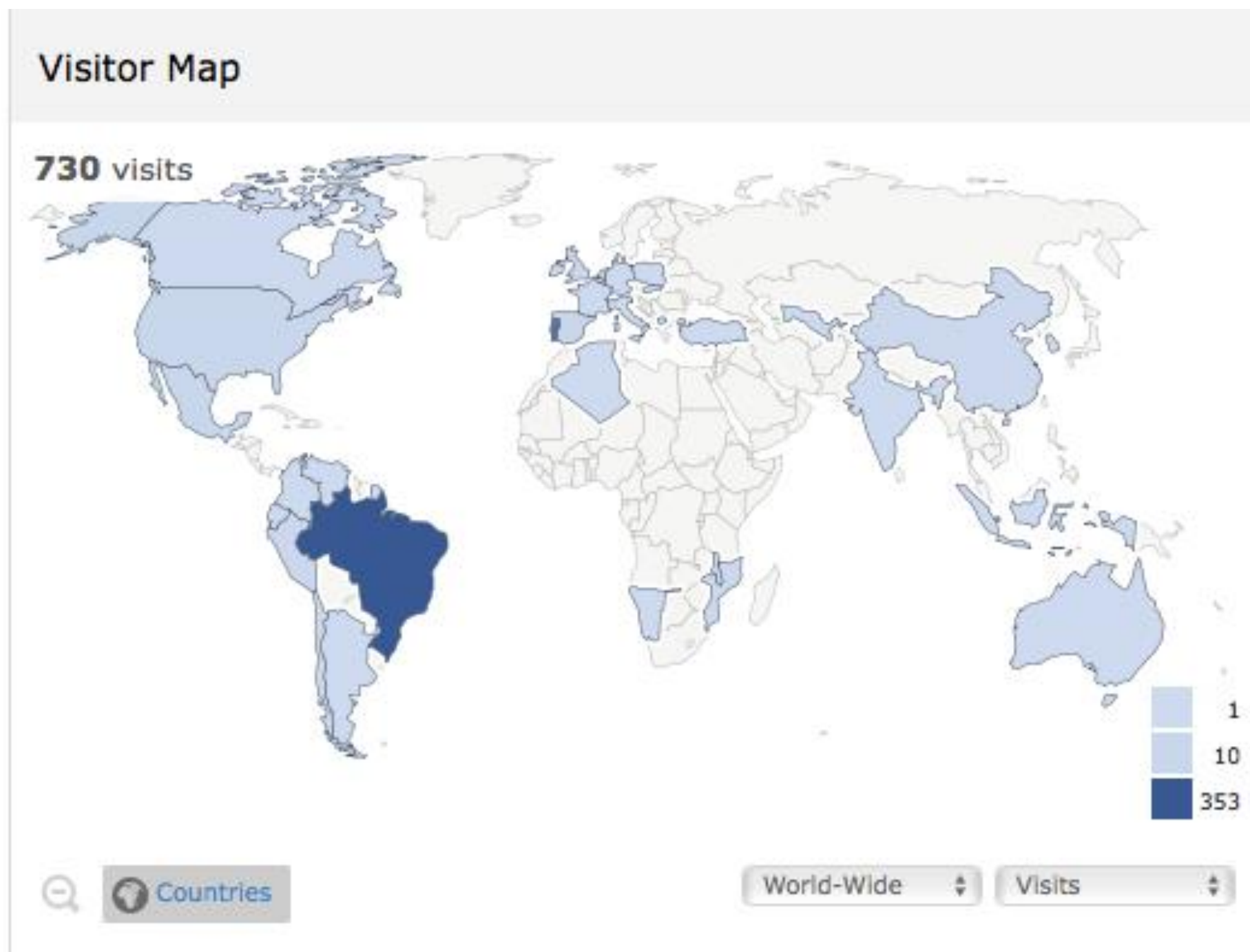
Usage Activity Statistics

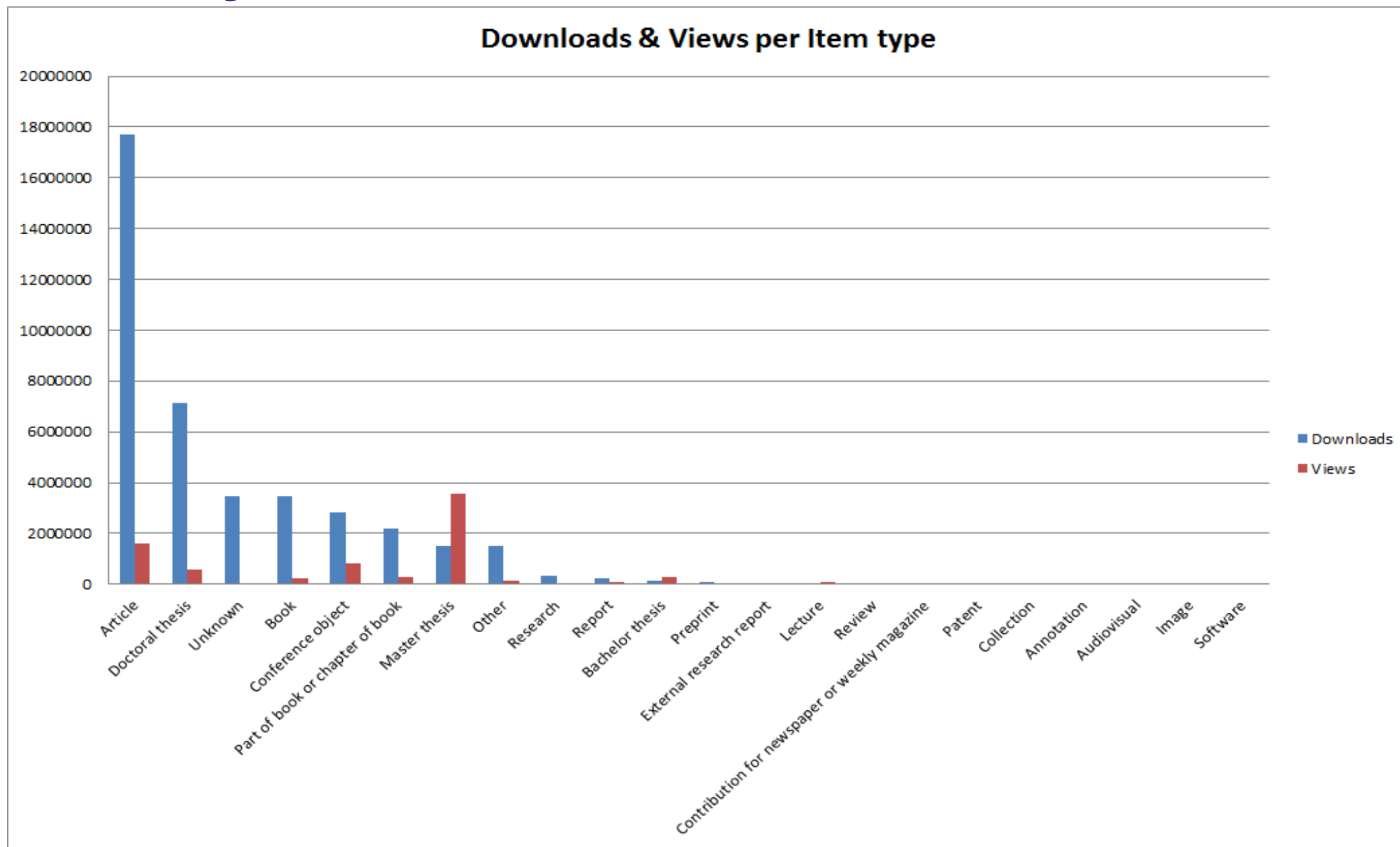**Number of Downloads**

Legend: 2016, 2017, 2018

Categories: IRUS-UK, OpenAIRE(Matomo), SARC-OJS

Usage Activity Statistics

Number of Views (only by Matomo)

Usage Activity Statistics

Repository Downloads

Usage Activity Statistics

Downloads & Views per Item type
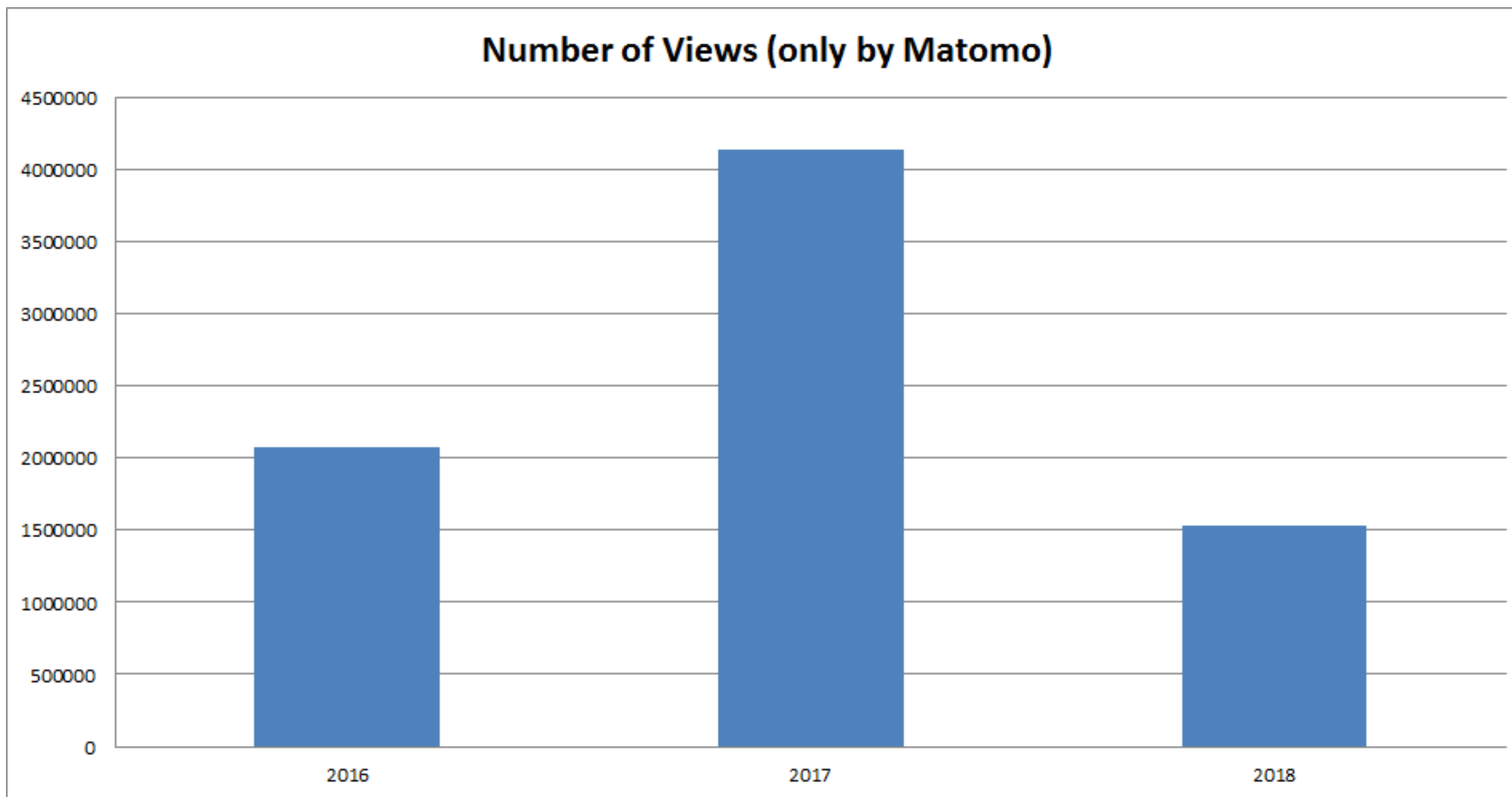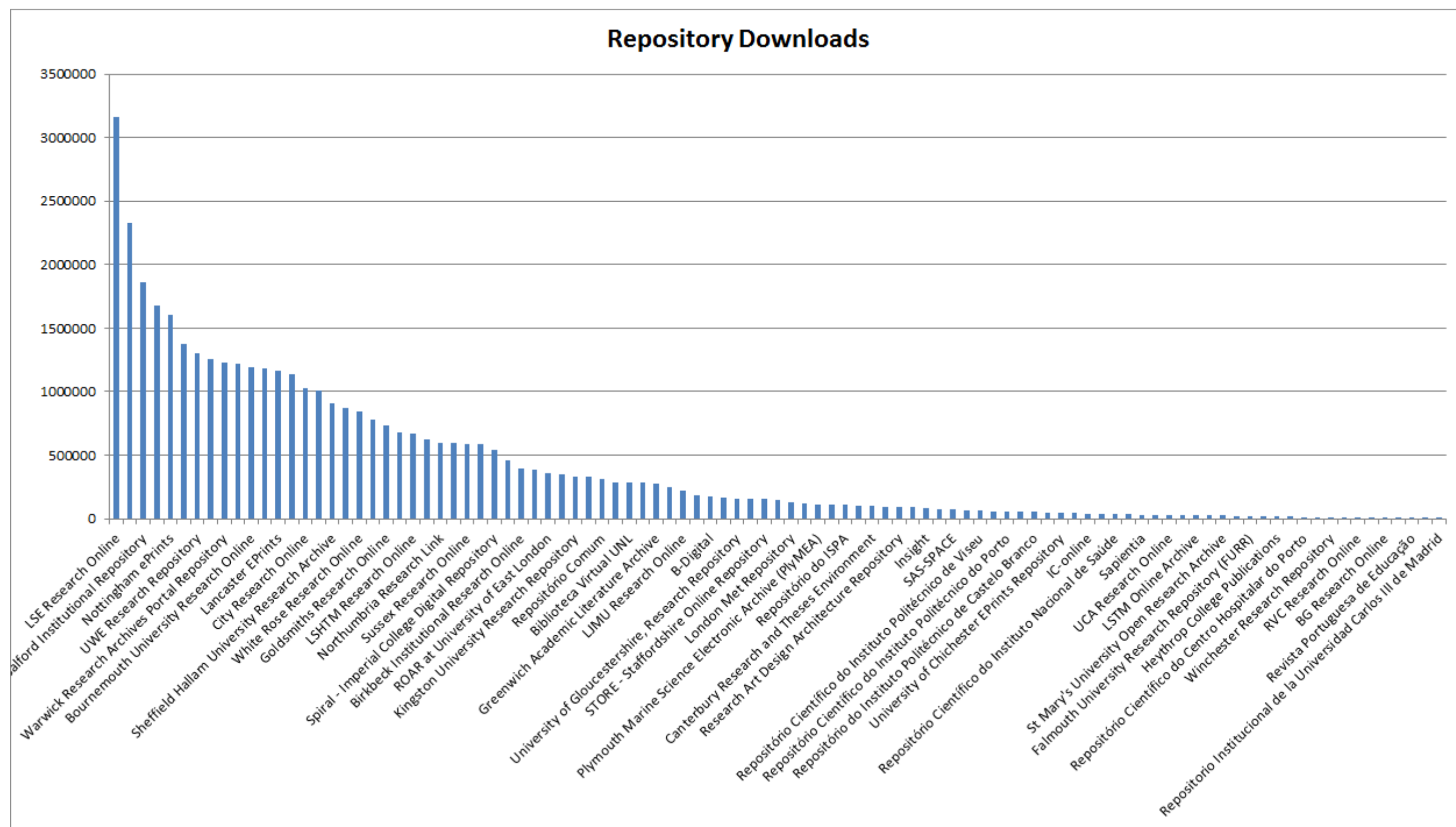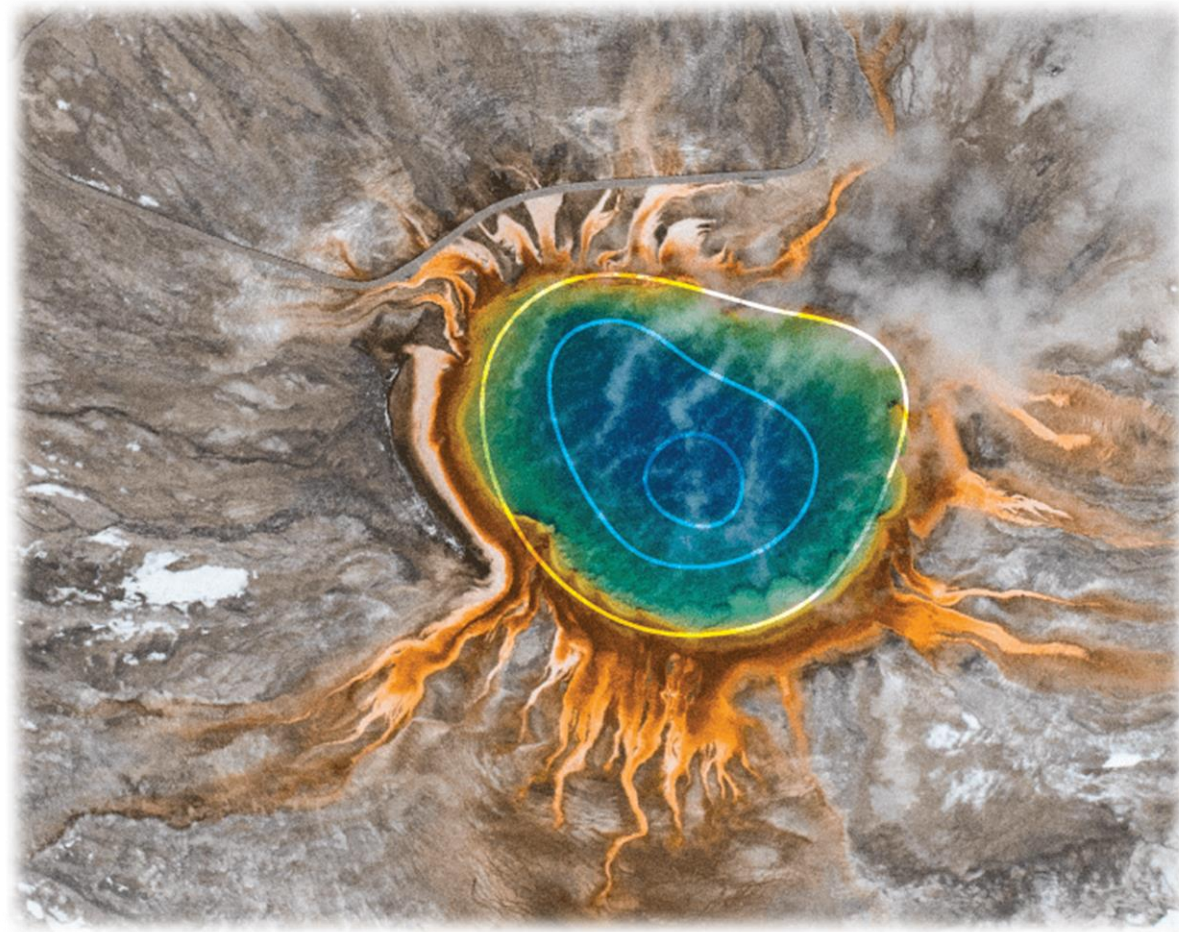
2

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

OpenAIRE | PROVIDE

Home   Sources   Compatibility   Content   Metrics
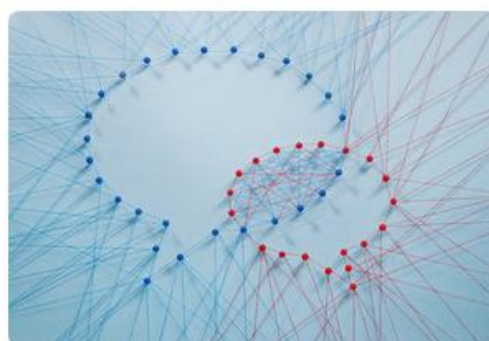
University Minho

Register

Register data sources in the OpenAIRE infrastructure

Validate

Validate data sources against OpenAIRE guidelines

Notifications

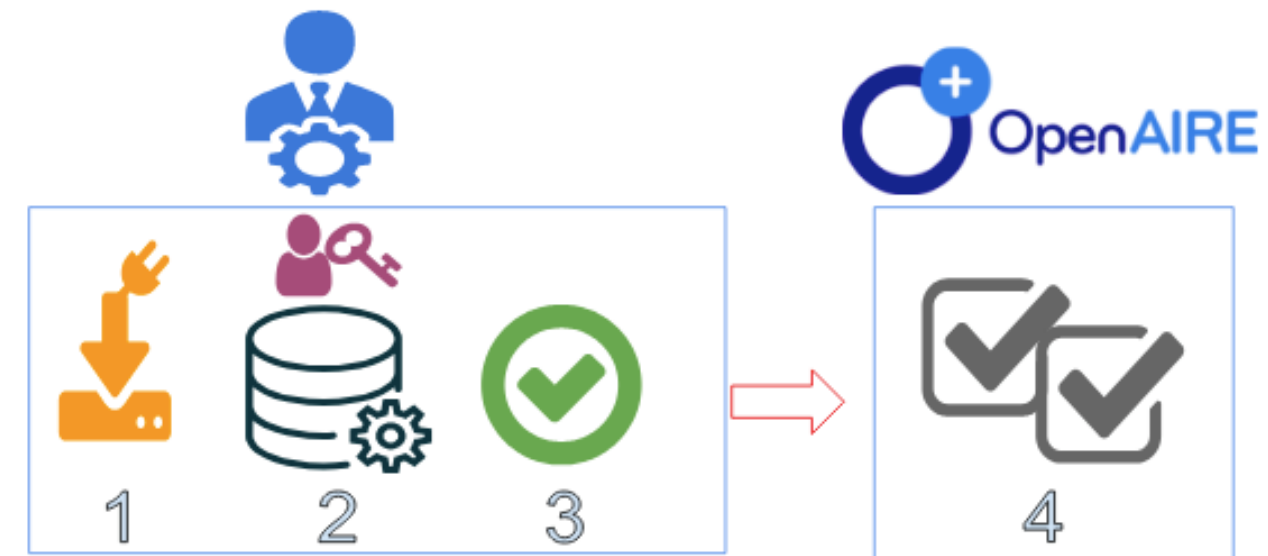View notifications to enrich the metadata and the content

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

Configure Metrics for selected Datasource

OpenAIRE | JOIN   Home   Sources   Compatibility   Content   Metrics   Admin   Dimitris Pierrakos | LOGO

Metrics Configuration & Software Details

Info / Tracking setup

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)
- as an EPrints plugin for version 3 (https://github.com/openaire/EPrints-OAPiwik)

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:

PiwikID
   000

AuthenticationToken
   01233456

Details for the configuration files are given in the README of the tracking code.

OpenAIRE

European Commission

Usage Metrics by example of the UMinho Repository

Usage Metrics on the Article Level

SUSHI-Lite Interface

- **Available as beta with the help of IRUS-UK**
  - 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.

Research Software Usage Report tbc.

SUSHI response example (JSON)

Repository Report	Item Report

```
- {
    - ItemIdentifier: [
        - {
            Type: "OpenAIRE",
            Value: "opendoar____::8e98d81f8217304975ccb23337bb5761"
        },
        - {
            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"
        },
```

```
- Report: {
    @Created: "2017-09-06 08:00:21+0000",
    @Version: "4",
    @Name: "IR1:4",
    - Vendor: {
        - Contact: {
            Contact: "OpenAIRE Helpdesk",
            E-mail: "helpdesk@openaire.eu"
        },
        Name: "OpenAIRE"
    },
    - Customer: {
        ID: "anonymous",
        - ReportItems: [
            - {
                - ItemIdentifier: [
                    - {
                        Type: "OpenAIRE",
                        Value: "dedup_wf_001::0233282d03f7f027b5c08890501849ef"
                    },
                    - {
                        Type: "URLs",
                        Value: "http://hdl.handle.net/1822/7975 ;http://hdl.handle.net/1822/7463 ;http://europepmc.org/articles/PMC2268319 ;"
                    },
                    - {
                        Type: "OAI",
                        Value: "oai:europepmc.org:1834183"
                    }
                ],
                ItemPublisher: "American Society for Microbiology",
                ItemPlatform: "Universidade do Minho: RepositoriUM",
                ItemDataType: "Article",
                ItemName: "Adaptive evolution of a lactose-consuming Saccharomyces cerevisiae recombinant",
                - ItemPerformance: [
                    - {
                        - Period: {
                            Begin: "2017-01-01",
                            End: "2017-01-31"
                        },
                        - Instance: [
                            - {
                                MetricType: "ft_total",
                                Count: "1"
                            },
                            - {
                                MetricType: "abstract",
                                Count: "4"
                            }
                        ],
                        Category: "Requests"
                    },
```

3

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)

OpenAIRE

European Commission

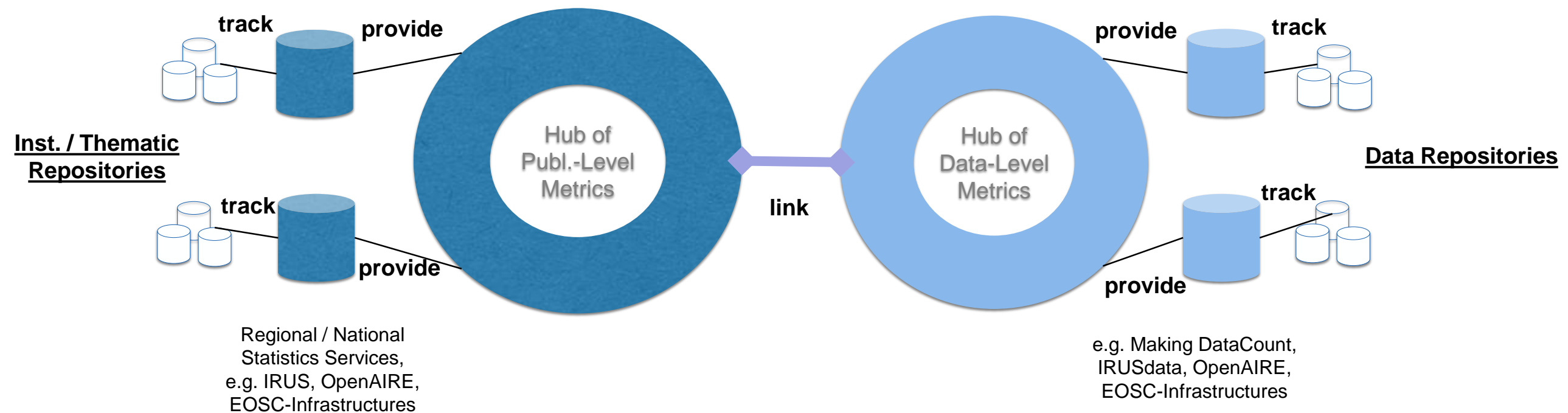"The larger the network the better the usage statistics"

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**

track **provide**

Inst. / Thematic Repositories

track **provide**

Hub of Publ.-Level Metrics

link

Hub of Data-Level Metrics

provide **track**

Data Repositories

provide **track**

Regional / National Statistics Services, e.g. IRUS, OpenAIRE, EOSC-Infrastructures

e.g. Making DataCount, IRUSdata, OpenAIRE, EOSC-Infrastructures

OpenAIRE

European Commission

👆 www.openaire.eu

🐦 @openaire_eu

f facebook.com/groups/openaire

andreas.czerniak@uni-bielefeld.de

✉ dpierrakos@gmail.com

jochen.schirrwagen@uni-bielefeld.de