

Scaling Usage Statistics across Repositories as an OpenAIRE Analytics Service

Dimitris Pierrakos, *ATHENA Research & Innovation Center*

Jochen Schirrwagen, *Bielefeld University*

Pedro Miguel Oliveira Bento Príncipe, *University of Minho*

Ricardo Saraiva, *University of Minho*



Outline

- Introduction
- Methodology
- Pilots & Preliminary Results
- Conclusions & The Future



Outline

- Introduction
- Methodology
- Pilots & Preliminary Results
- Conclusions & The Future



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 analytics service for usage data collected from data providers



Usage Analysis Service: Aims

- Standard alignment across heterogeneous data providers for gathering usage data & sharing statistics.
- Taking care of data privacy policies in EU and member states.
- Collection, measure and analysis of usage data (downloads and views).
- Correlate with other altmetrics.



Outline

- Introduction
- **Methodology**
- Pilots & Preliminary Results
- Conclusions & The Future



Methodology

- Tracking phase
 - Tier 1 approach: direct tracking
 - Tier 2 approach: exploit Sushi Lite API
- Analysis phase
 - Import
 - Process (COUNTER4 compliance)
 - Analyze



Using Piwik in OpenAIRE

- An Open Source analytics platform
- Tracking via JavaScript embedded in Web pages
- Usage parameters: VisitorID, SessionID, Visitor IP, Timestamp, Country, and many more
- IP anonymization enabled
- Bots handling

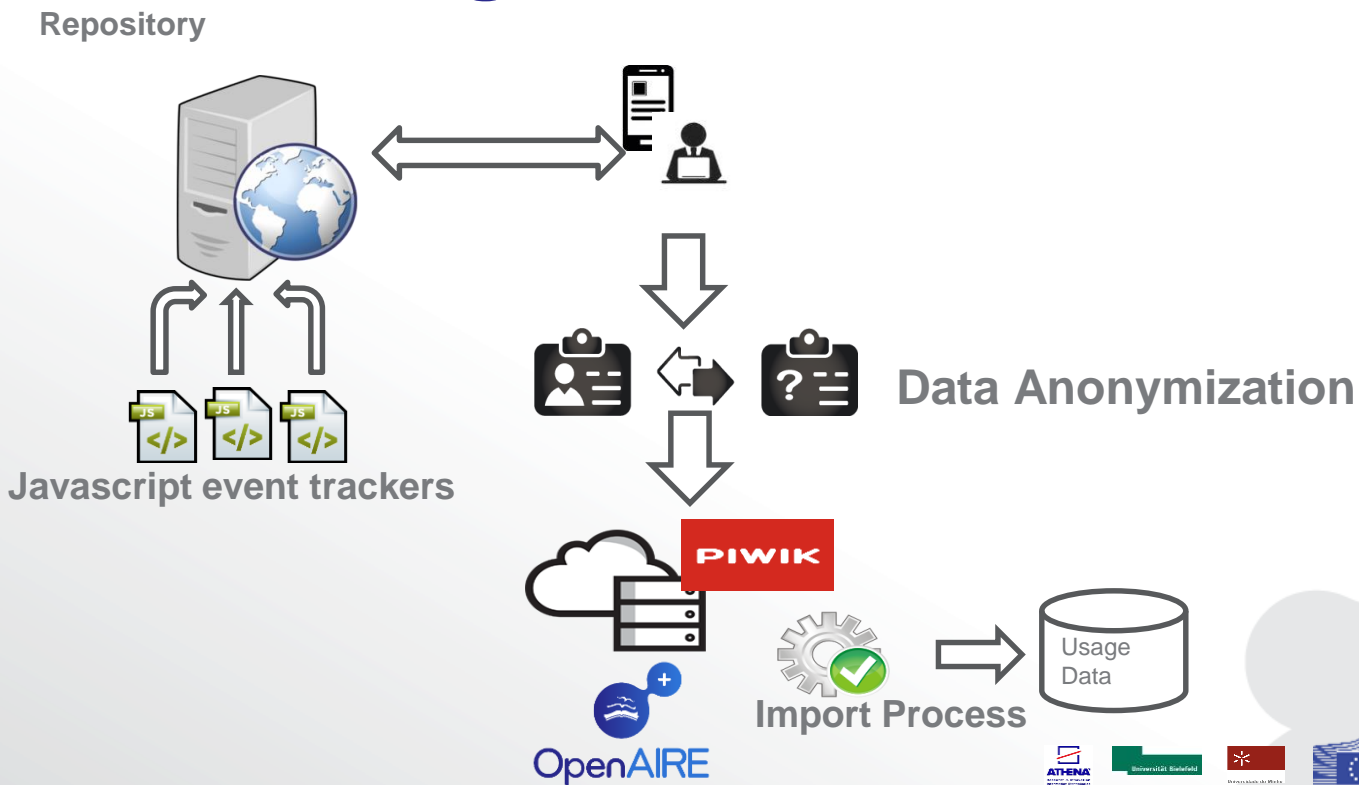


Counter Code of Practice

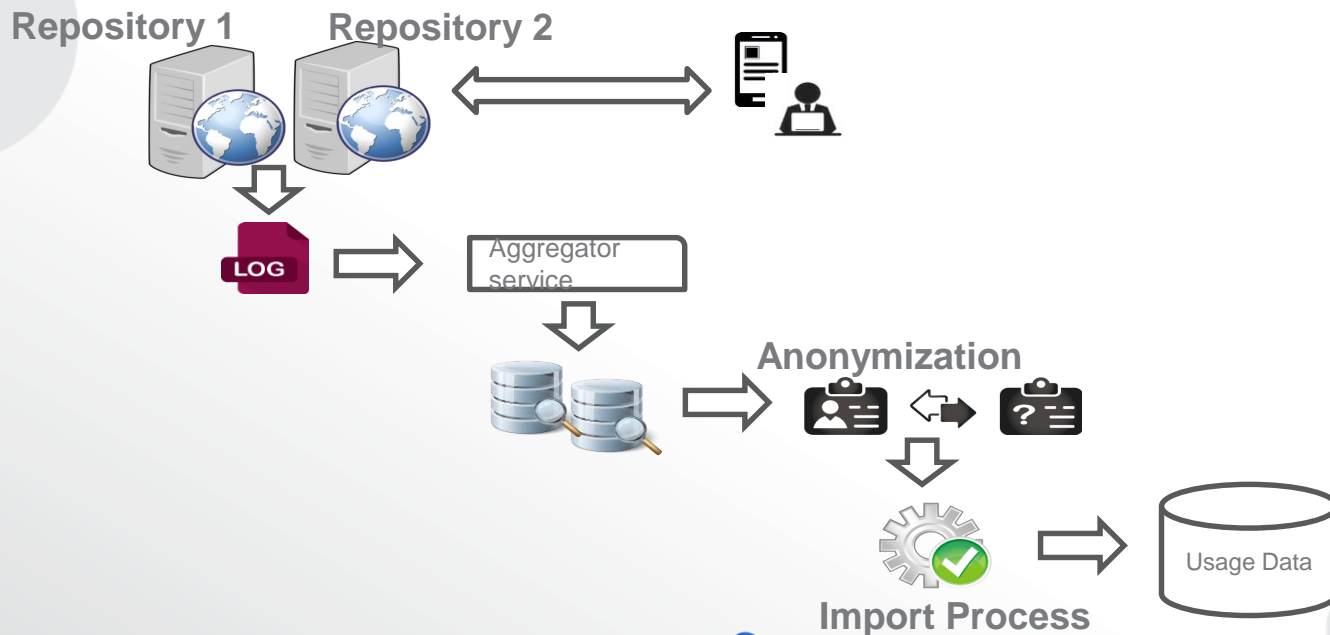
- An International, extendible Code of Practice for e-Resources.
- Measures usage information in a credible, consistent and compatible way using vendor-generated data.
- Specifications for:
 - Data Collection & Processing
 - Usage Analysis Reports
- **Currently in Release 4**



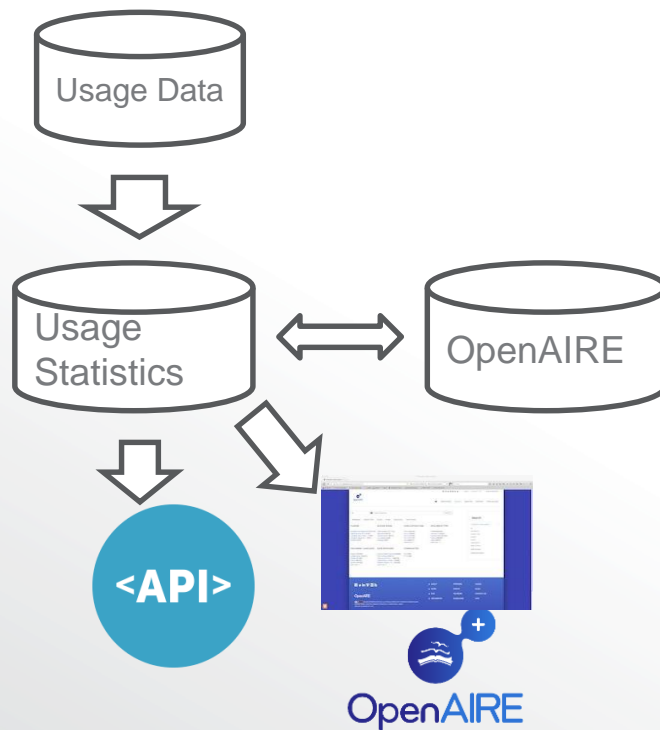
Tier 1 Tracking Workflow



Tier 2: Aggregated Statistics Workflow using SUSHI Lite

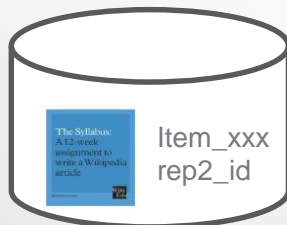
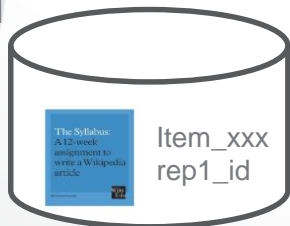


Usage Data Analysis

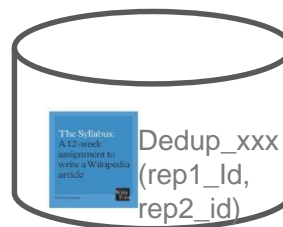


Deduplication Process

Repository 1



Deduplication



Repository 2

- Enhances the calculation of usage statistics by having a single id for common records
- Disseminate cross-repository usage statistics

Outline

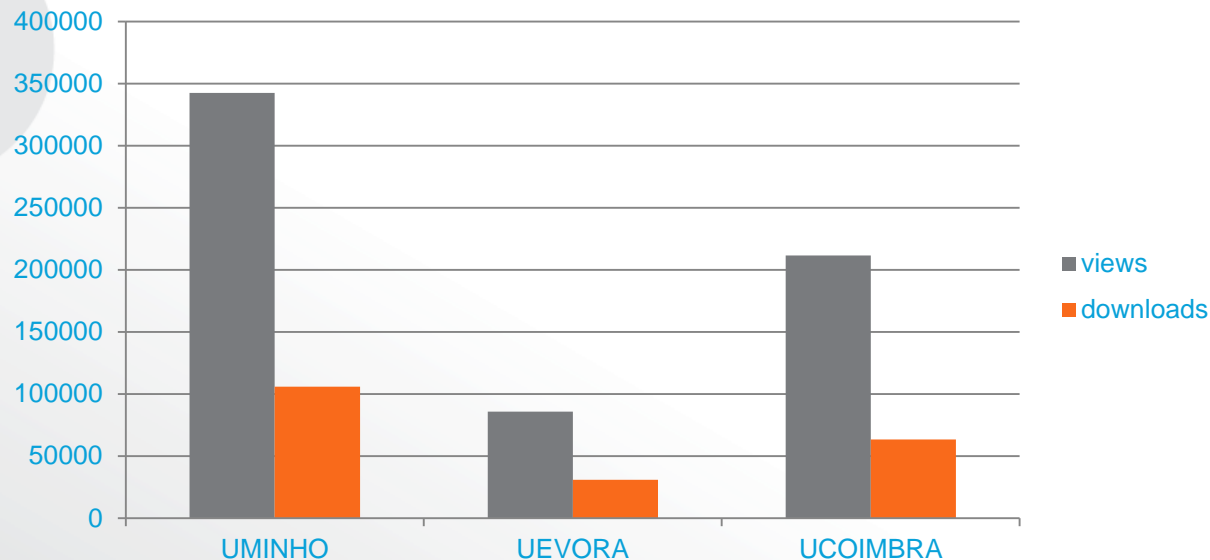
- Introduction
- Methodology
- **Pilots & Preliminary Results**
- Conclusions & The Future



Pilots 2nd phase



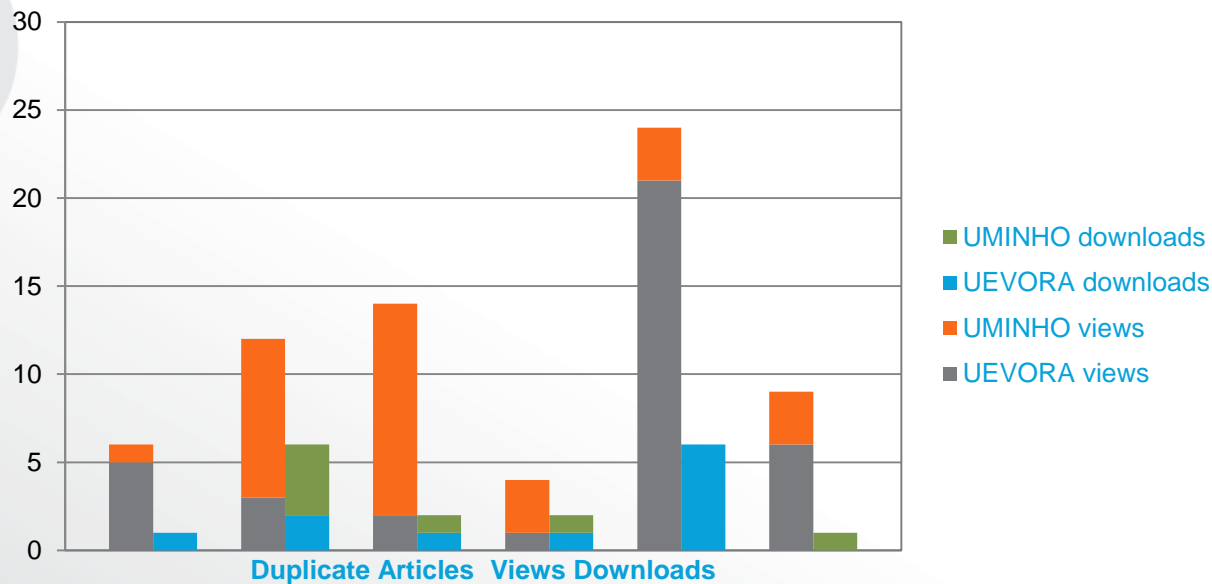
Preliminary Results



Metadata Views – Downloads on Pilot Repositories



Preliminary Results



Metadata Views – Downloads Duplicate Information



Outline

- Introduction
- Methodology
- Pilots & Preliminary Results
- **Conclusions & The Future**



Conclusions

- ✓ Usage Analysis in OpenAIRE
- ✓ Methodology
- ✓ Pilot Results
- ✓ Challenges
 - ✓ Better handling of bots tracking and “gaming” activity in usage data
 - ✓ Tackling of direct downloads



The Future

- ✓ **Collaboration with National Open Access Desks (NOADs) for usage service dissemination**
- ✓ **Beta release in 2016**
- ✓ **Production release in 2017.**





<http://openaire.eu>



@openaire_eu



<https://www.facebook.com/groups/openaire/>



<https://www.linkedin.com/groups/3893548/profile>



info@openaire.eu

