

Why we need a Network of Usage Data Providers - OpenAIRE Impact Metrics Results

Jochen Schirrwagen

Bielefeld University

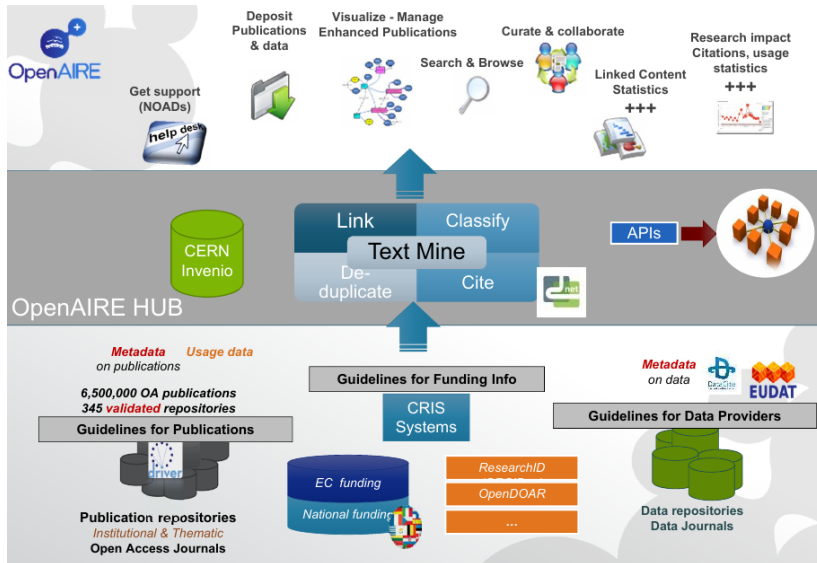
- 1 **OpenAIRE EU-FP7 project**
- 2 **Workpackage on Impact Metrics**
- 3 **Usage Statistics Service**
- 4 **PlosOpenR Use Case**
- 5 **Implementation Overview**
- 6 **Usage Statistics Services - Challenges**
- 7 **Vision**

Background

- Project period 12/2009 - 11/2012
- Making FP7 / ERC research output visible
- OA-pilot implementation (SC39)
- OpenAIREplus follow-up project (12/2011 - 05/2014)
 - Merge DRIVER & OpenAIRE : Knowledge Infrastructure
 - Extend scope on beyond FP7; national funders
 - enriched links: publications - research data - projects
- Current figures (04/2013)
 - FP7 projects / in SC39: 19.097 / 1448
 - bibliographic records on FP7 articles: > 41.000
 - Open Access ratio approx. 30%
 - registered repositories: 130
 - text-mining on FP7 acknowledgment and claiming of articles

OpenAIRE Knowledge Infrastructure

• <http://www.openaire.eu>



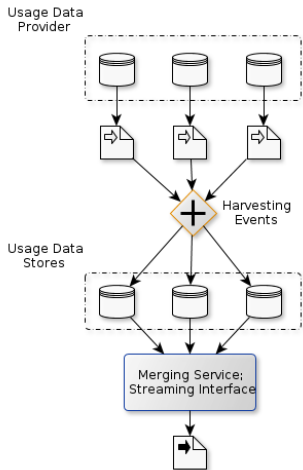
- monitoring of FP7 SC39 research output
- definition of methodologies to measure impact
- calculating quantitative indicators
 - Bibliometrics
 - Webometrics
 - supplement with Usage Statistics

Group	Indicator	Source
Usage Stats	Views	Log files
	Downloads	Log files
Webometrics	URL mentions	Search engines
	Title (DOI) mentions	Search engines / Social Networks
	Readings	Readings in Mendeley
Bibliometrics	Citations	WoS, Scopus, Google Scholar

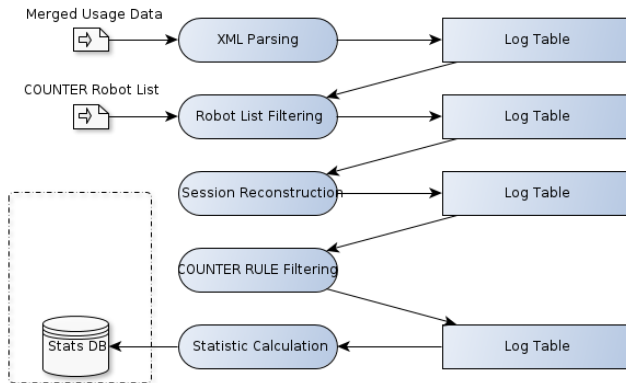
- limited by quantity of article corpus (FP7, SC39)
- limited by time window (FP7: 2007 - 2013) vs. citation window
- web analysis tools from commercial search engines disappeared
 - e.g. Yahoo SiteExplorer
- limited by no. of participating repositories registered in OpenAIRE
- need for article de-duplication / statistics consolidation
- only few articles minted with persistent identifiers

- 1 Guidelines for Transfer Protocol and Format
 - aligned with KE-USG
- 2 Indicator Specification
- 3 Usage Statistics Service
- 4 Reporting to the EC
 - Methodology
 - Impact Metrics Results

Usage Data Collection



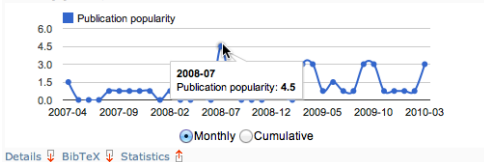
Processing and Calculation



Article Popularity

Key management and link-layer security of wireless sensor networks : Energy-efficient attack and defense

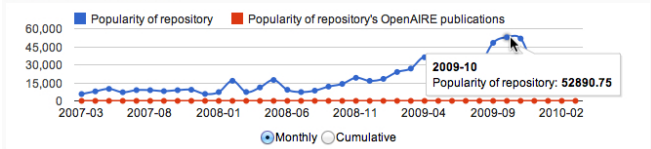
Author(s): Law, Y.W.



Repository Popularity

Universiteit Twente Repository

Show details Hide statistics



- thematic and institutional repositories were approached
 - many repository managers interested but with hesitation
 - ! effort needs to be calculated
 - ! policy about the use of the usage data & statistics
 - ! service sustainability vs. project & pilot character
- participating repositories: CERN-Physics, GoeScholar, OpenAIRE-Orphan, PUB repository, Twente repository, Utrecht repository

- using madIS framework
- usage data from Dutch repositories
- processed on Core2 Duo 3 Ghz, 4 GB workstation

Process	# Events after	elapsed Time	Rate
Download to local store	7,481,937	13 min	16.9 MB/sec
XML parsing	7,481,937	54 min 39 sec	4 MB/sec
Robot filtering	6,630,121	11 min 16 sec	
Session reconstr. (2,277,305)	6,630,121	20 min	
COUNTER rule filtering	4,510,527	5 min 30 sec	
Statistics calc.	4,510,527	5 min 41 sec	
total time		1 hour 50 min	

- Collaboration between PLOS-ALM and OpenAIRE
- Development of tools exploring PLOS-ALM on FP7 funded articles
- Experimentation with visualizations

Data Sources

- PLOS Search API; PLOS-ALM API
- OpenAIRE OAI on FP7 project information

Method

- PLOS search on financial disclosure field
- Text-Mining, Normalisation, Filtering
- Aggregation of relative ALM data
- Data Exploration & Visualization

Result

- 1166 PLOS papers acknowledging 624 FP7 projects

- What to consider if combining indicators of different metrics ?
- Issues
 - Does each metric cover the same corpus of articles ?
 - How to specify the optimal weighting of each indicator ?
 - indicator peaks in social networks vs. long term citation counts
- Example for Illustration
 - PLOS article "The Network of Global Corporate Control"
 - published: 26-Oct-2011, doi: 10.1371/journal.pone.0025995
 - reception of science in media

WIN THE TRIP OF A LIFETIME! [Click here.](#)

Forbes ▾

New Posts

+17 posts this hour

Most Popular

The Facebook Job Board

Lists

100 Largest Charities

Video

Petraeus' Digi...



Bruce Upbin, Forbes Staff

I manage our tech and wealth teams.

+ Follow (422)

Subscribe

3.7k

TECH | 10/22/2011 @ 9:37AM | 344,985 views

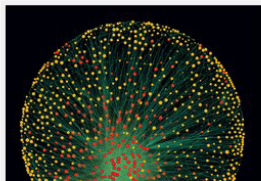
The 147 Companies That Control Everything



67 comments, 60 called-out

+ Comment now

Three systems theorists at the Swiss Federal Institute of Technology in Zurich have taken a database listing 37 million companies and investors worldwide and analyzed all 43,060 transnational corporations and share ownerships linking them. [They built a model of who owns what and what their revenues are](#)



Mo

NEV

Wl

Th

+45

Pe

Th

Jo

Ov

Th

Mi

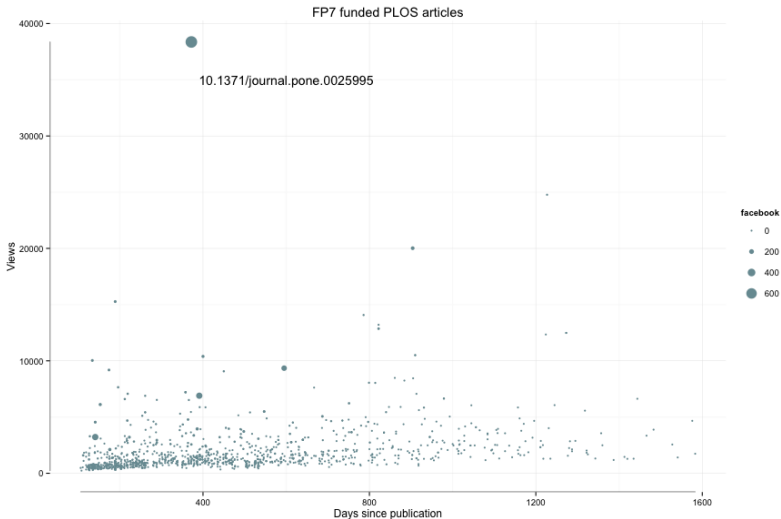
Hc

Ma

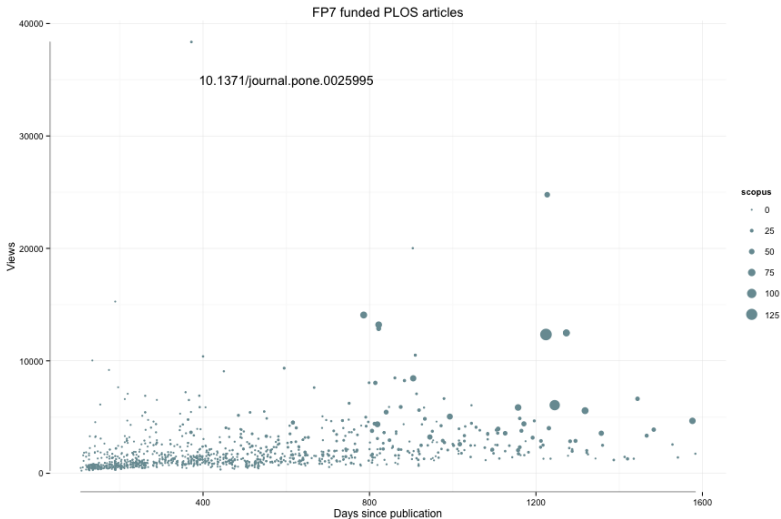
Hc

Th

Facebook Mentions (and in relation to other PLOS articles)



Scopus Citation Counts (and in relation to other PLOS articles)



- Aggregation: OpenAIRE D-NET, (Java),
<http://www.d-net.research-infrastructures.eu/>
- Usage Statistics: madIS (Python, using SQLite),
<http://code.google.com/p/madis/>
- PlosOpenR: (R),
<https://github.com/articlemetrics/plosOpenR>

- National Initiatives on Usage Statistics
 - IRUS-UK, SURFSure, OA-Statistics, LogEc
- OpenAIRE Usage Statistics
- Repository Networks, e.g . DiVA using AwStats
- Statistics Plugins for repository platforms

Main Barriers on the Repository Manager Level

- effort and support needed
- hesitation of repository managers
- lack of standardized, well accepted indicators

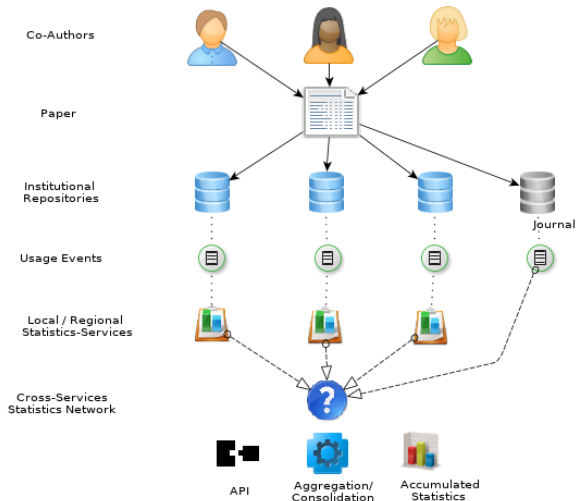
Organizational Barriers on the Serviceprovider Level

- data exchange bi-directional with altmetrics services
- event-type considered: "download only" or "metadata view and download hit"
- comparing statistics results

Technical Barriers on the Serviceprovider Level

- bottleneck OAI-PMH as not designed for large bulk data transfer
- different interpretation / use of ContextObjects describing Usage Events
- different level of filtering due to lack of shared and maintained robot list

Scenario for a Usage Statistics Network



Contact

✉ jochen.schirrwagen@uni-bielefeld.de

● <http://www.openaire.eu>

Contributors to the Workpackage

- CSIC: I.F.Aguillo, S.P.Álvarez
- CERN: S. Kaplun, L.H.Nielsen
- NKUA: N. Manola, L. Stamatogiannakis, M. Triantafillidi, M. Vayanou,
- SURF: Maurice Vanderfeesten
- UNIBI: W. Horstmann, N. Jahn, M. Imialek, M. Loesch

Authors and Contributors to PlosOpenR

- N.Jahn (UNIBI), M.Fenner (PLOS-ALM), H. Dimitropoulos (NKUA)