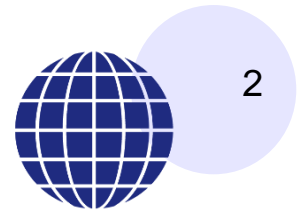


Impact Metrics Services and Repositories as Usage Data Providers

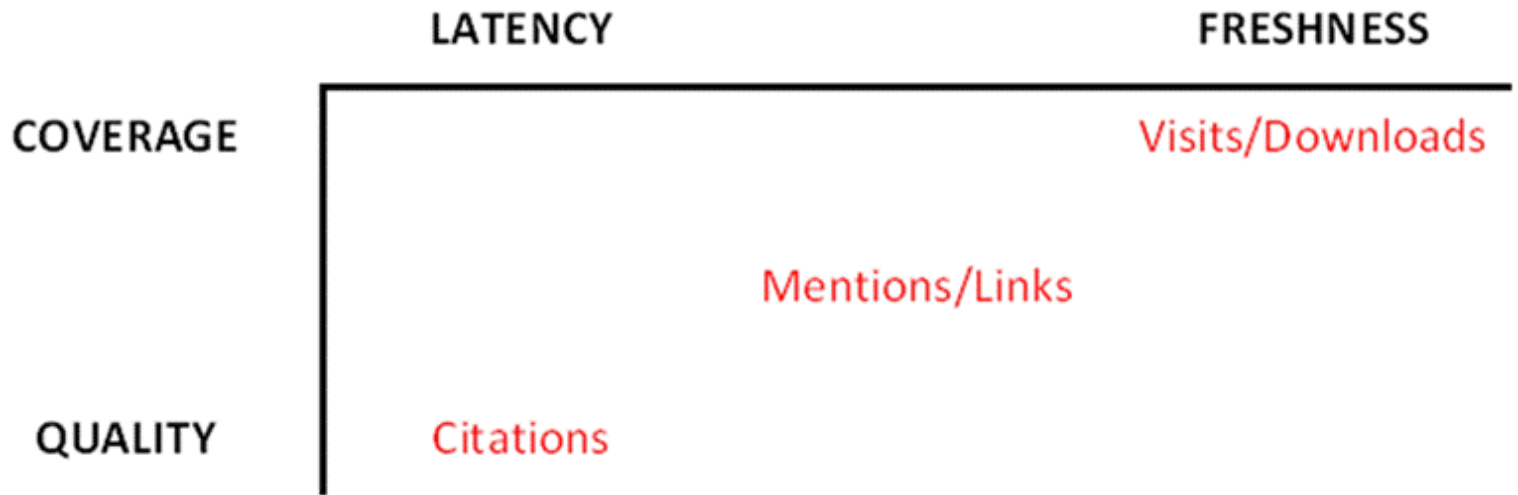
Jochen Schirrwagen
Istanbul, May 08 2013
COAR Annual Assembly

Agenda

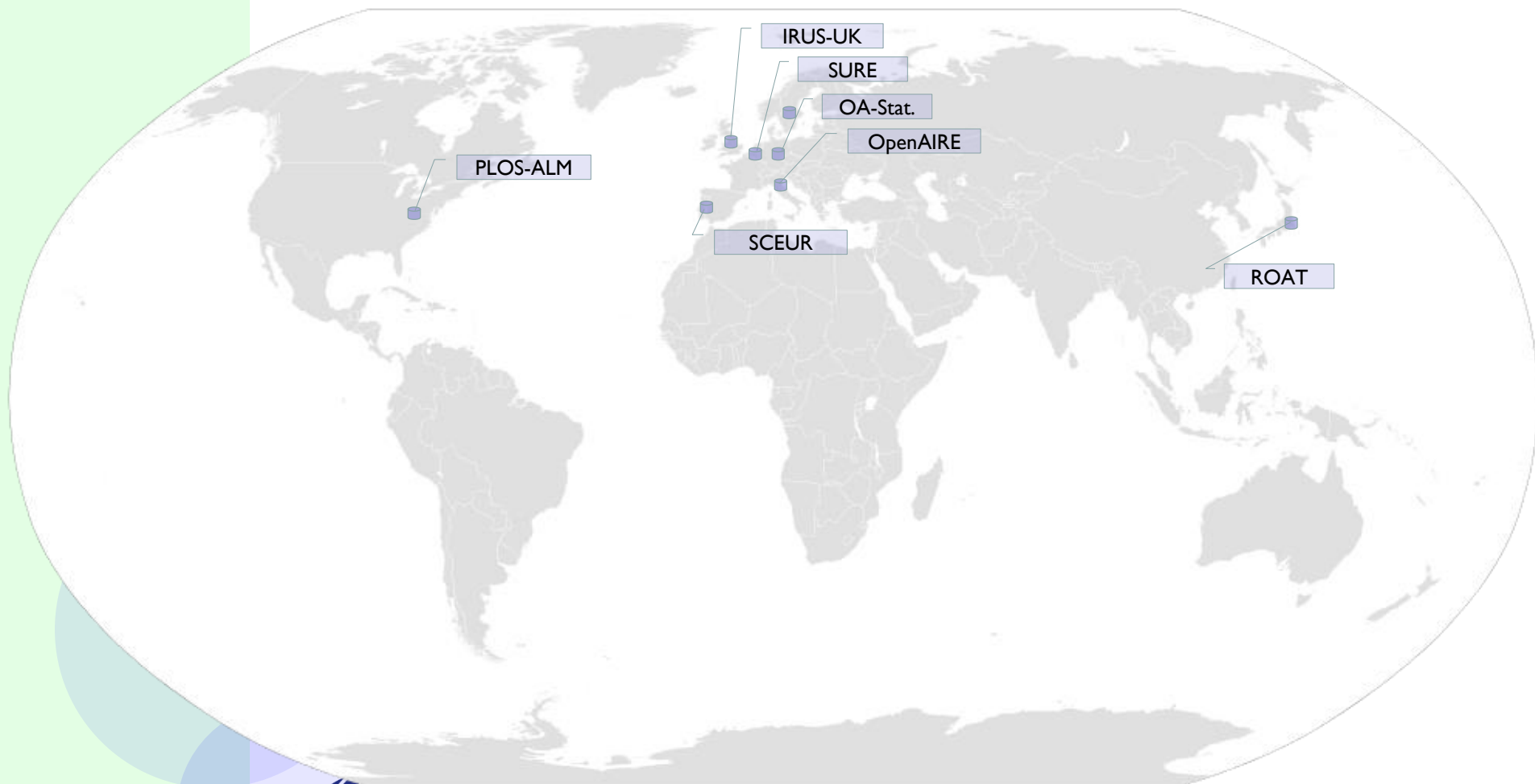
- Objectives of the tutorial
- Overview of statistics and metrics services
- Usage Data and Statistics Workflow
- Calculation
- Examples



Indicators and Metrics of Impact



Statistics and Metrics Services



Impact Metrics vs. Usage Statistics

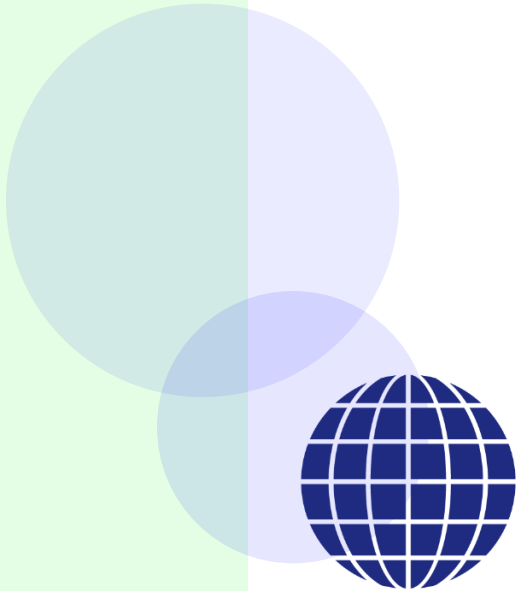
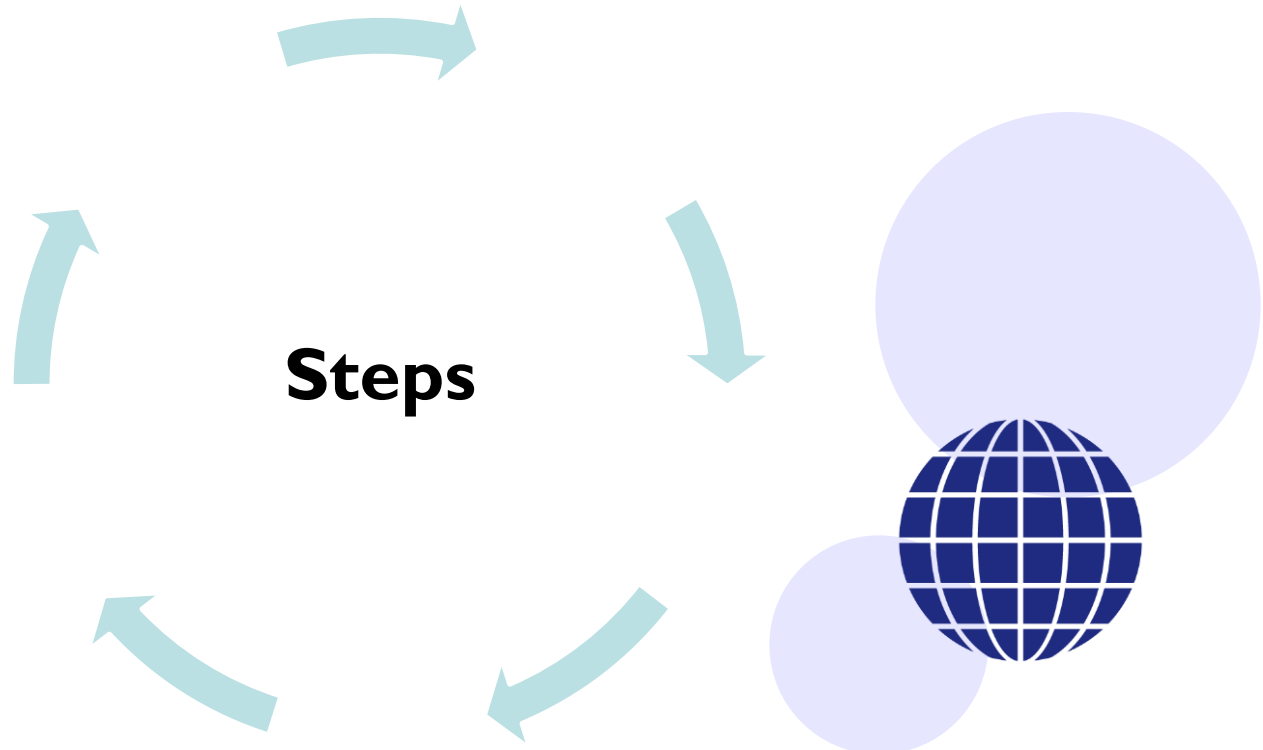
- AltMetrics services
 - ImpactStory
 - ReaderMeter
 - ScienceCard
 - PLUMAnalytics
 - PaperCritic
 - Crowdometer: tweets linking to research articles, with users adding semantic informatin based on cito-ontology



Webometrics - Overview

	Altmetrics	Usage Statistics
Indicators	Tweets, likes,bookmarks, readers, mentions	Downloads, views
Audience	Academic & Public	Rather academic
Resource	Research paper; Refs. in other media	Research paper; repository assets
“Speed”	Immediately	Immediately
Coverage	Datasources used by altmetrics-service	Often limited to individual repository
Noise/bias	“gaming”	robots





#1 - User initiated Access

- User or robot access a resource provided by a repository
- Direct access
- Or directed access



#11 – Record Access in Log-File

- Log entry typically recorded as NCSA combined log format



Anatomy of a Log File Entry

- request protocol, method, datestamp, path
- Response protocol code, size
- User agent, user ip address
- LogFormat "%h %l %u %t \"%r\" %>s %b
\"%{Referer}i\" \"%{User-agent}i\" combined
- 127.0.0.1 user-identifier frank [10/Oct/2000:13:55:36 -0700]
"GET /apache_pb.gif HTTP/1.0" 200 2326



#III – Pseudonimization of IP-Address

- Data privacy
- Provide a salted hash for IP-Address
 - And/or for C-Class



#IV – Identification of Document Access from Logs

- E.g. mapping
 - requested filename to oai-record-id



#V – Transformation Logdata to Usage Data Format

- Using a common format to transfer usage data
 - Limited to information providing „context“ of the usage event



Metadata Format for Usage Events

- ContextObject Core Set
 - <referent>
 - <referringEntity>
 - <requester>
 - <service-type>
 - <resolver> and <referrer>



#VI – Usage Data Transfer

- FTP, e.g for raw log data dumps
- OAI-PMH
 - Pull-approach
 - Can be fetched incrementally
 - Not designed for large data amounts
- SUSHI
 - Automated consolidation of large amounts of usage data from different sources



#VI – Usage Data Transfer

- Tracker-Code

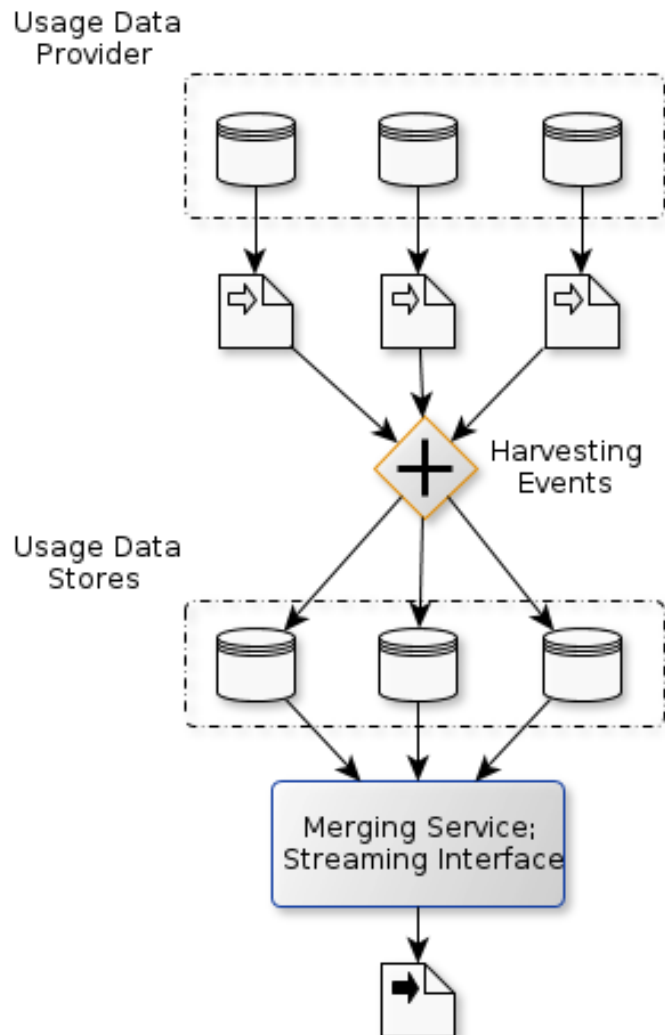
- OpenURL key/value pairs

- /tracker/?url_ver=Z39.88-2004&req_id=e02db545fbefd7d19bf24302a57f93ac&req_dat=Mozilla%2F5.0+%28compatible%3B+Googlebot%2F2.1%3B+%2Bhttp%3A%2F%2Fwww.google.com%2Fbot.html%29&rft.artnum=http%3A%2F%2Fspace.lib.cranfield.ac.uk%2Fhandle%2F1826%2F3228&svc_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Adc&svc.format=application%2Fpdf&svc_dat=Unknown&rfr_id=space.lib.cranfield.ac.uk&url_tim=2010-10-17T03%3A04%3A42Z&rft_id=info%3Adoi%3Ahttp%3A%2F%2Fdx.doi.org%2F10.1016%2Fj.istr.2008.10.006

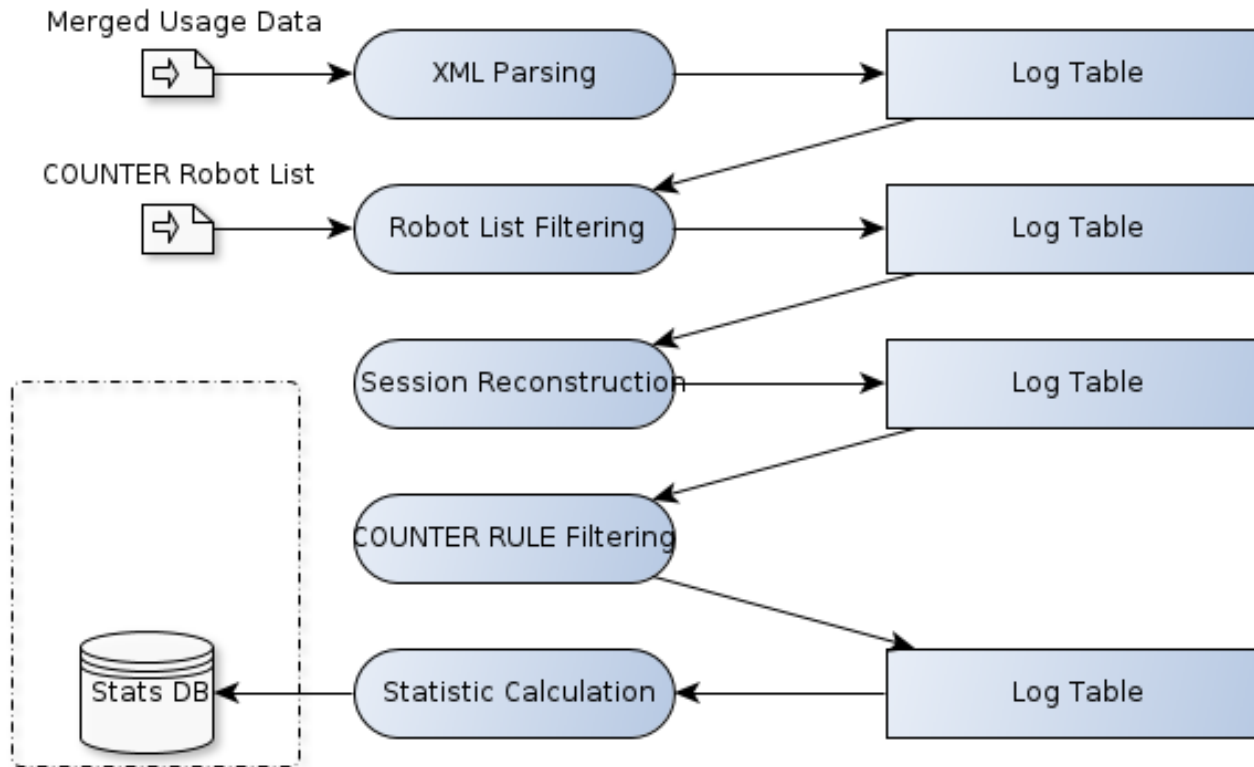
- See Google-Analytics



#VII – Usage Data Aggregation



#VII – Usage Data Aggregation



#VIII – Cleaning, Normalization

- Identification of robots by e.g. user-agent, IP-address range
- Sometimes ignoring robot.txt
- Robot lists
 - <http://purl.org/robotslist/current/robotlist.xml>
 - http://www.projectcounter.org/r4/COUNTER_robot_txt_list_Jan_2011.txt
- Merge of Double-Clicks as single user session

#VIII – Cleaning, Normalization

- Merge of Double-Clicks as single user session

Standard	Session duration
COUNTER	10 sec (HTML), 30 sec (PDF)
LogEc	1 month
AWStats	1 hour
IFABC	30 minutes



#IX – Usage Statistics Calculation

- Consolidation
- Delivery to the repository
 - csv, json, javascript-code in iframe



#X – Usage Statistics Presentation

- Indicate use of research result in the community
- Stimulate the access to research results
- On the level of the individual repository
- On the level of the author's publication list ?
- On the aggregator level
 - API ?
 - Reports ?
 - Dashboards ?

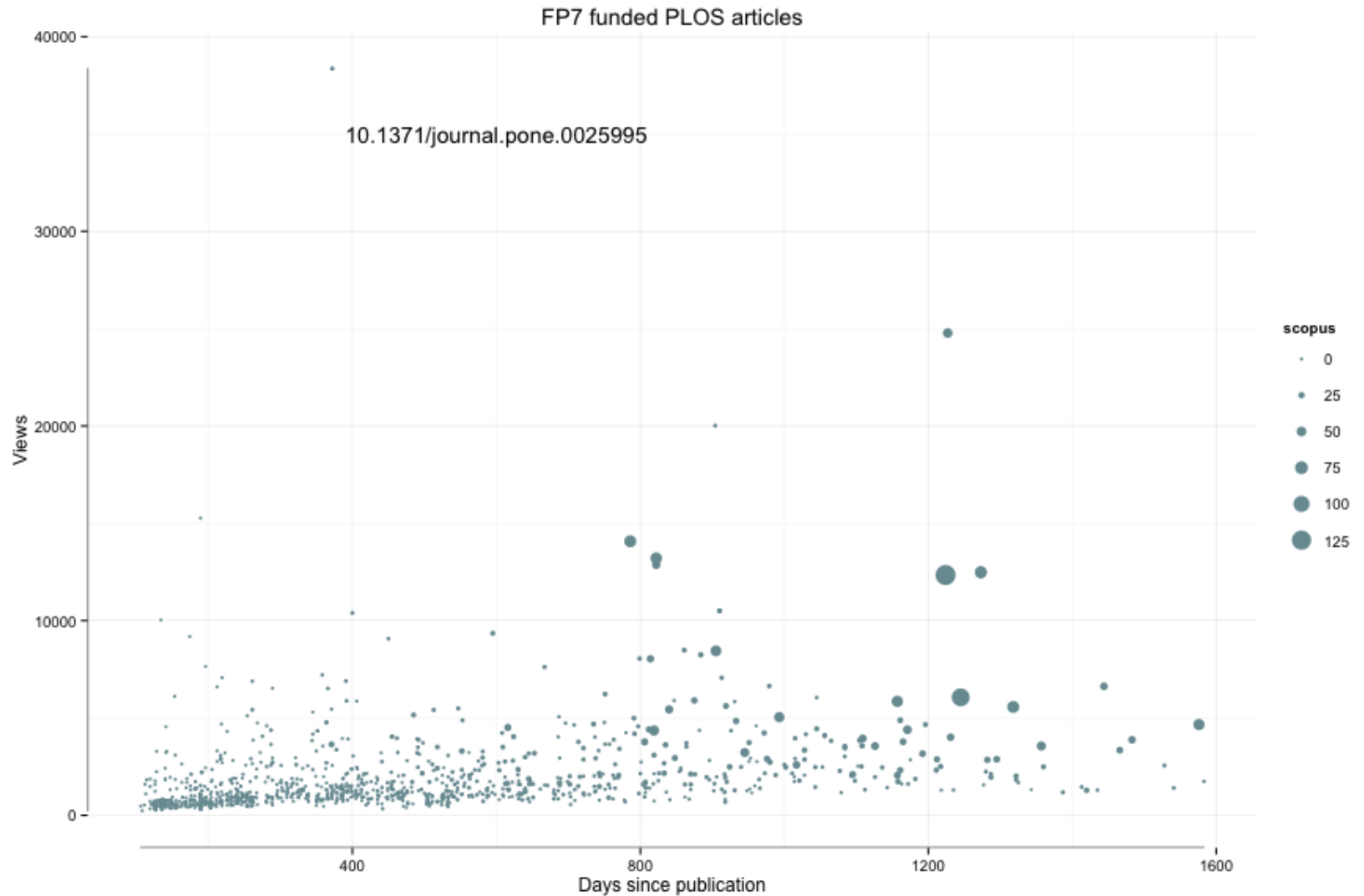


Issues

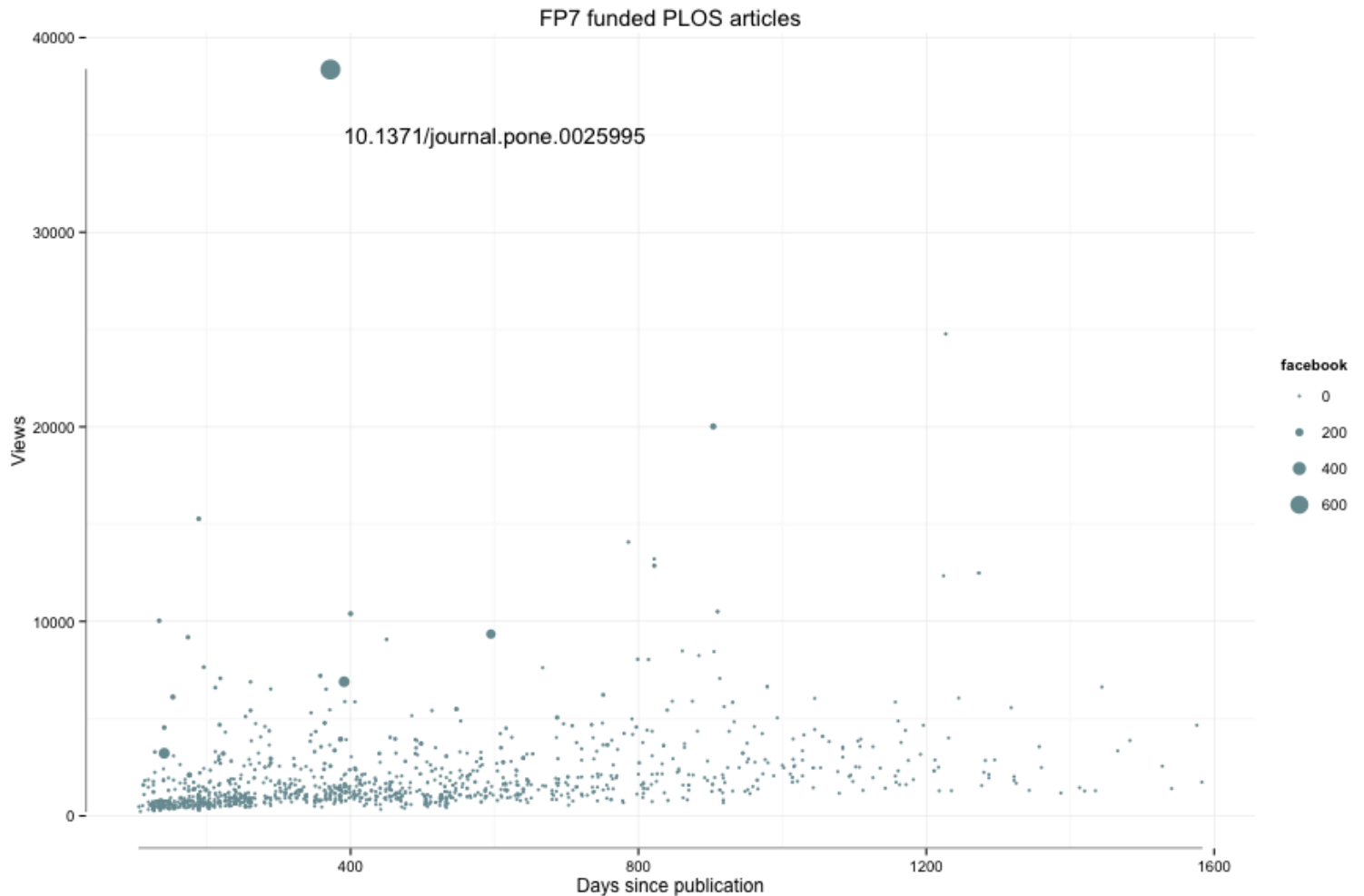
- Common robot list – shared and maintained by statistics services
- Publication: de-duplication ?
- Author: de-duplication ?
- Complex Objects – How to count ?



Correlation Citations – Social Media



Correlation Citations – Social Media Mentions



References

- <http://wiki.surf.nl/display/standards/KE+Usage+Statistics+Guidelines>
- <http://www.openaire.eu/en/support/guides/repository-managers/usage-stats>
- <http://www.dini.de/projekte/oa-statistik/english/>
- <http://www.irus.mimas.ac.uk/>



- Anything additional to fill in?!

