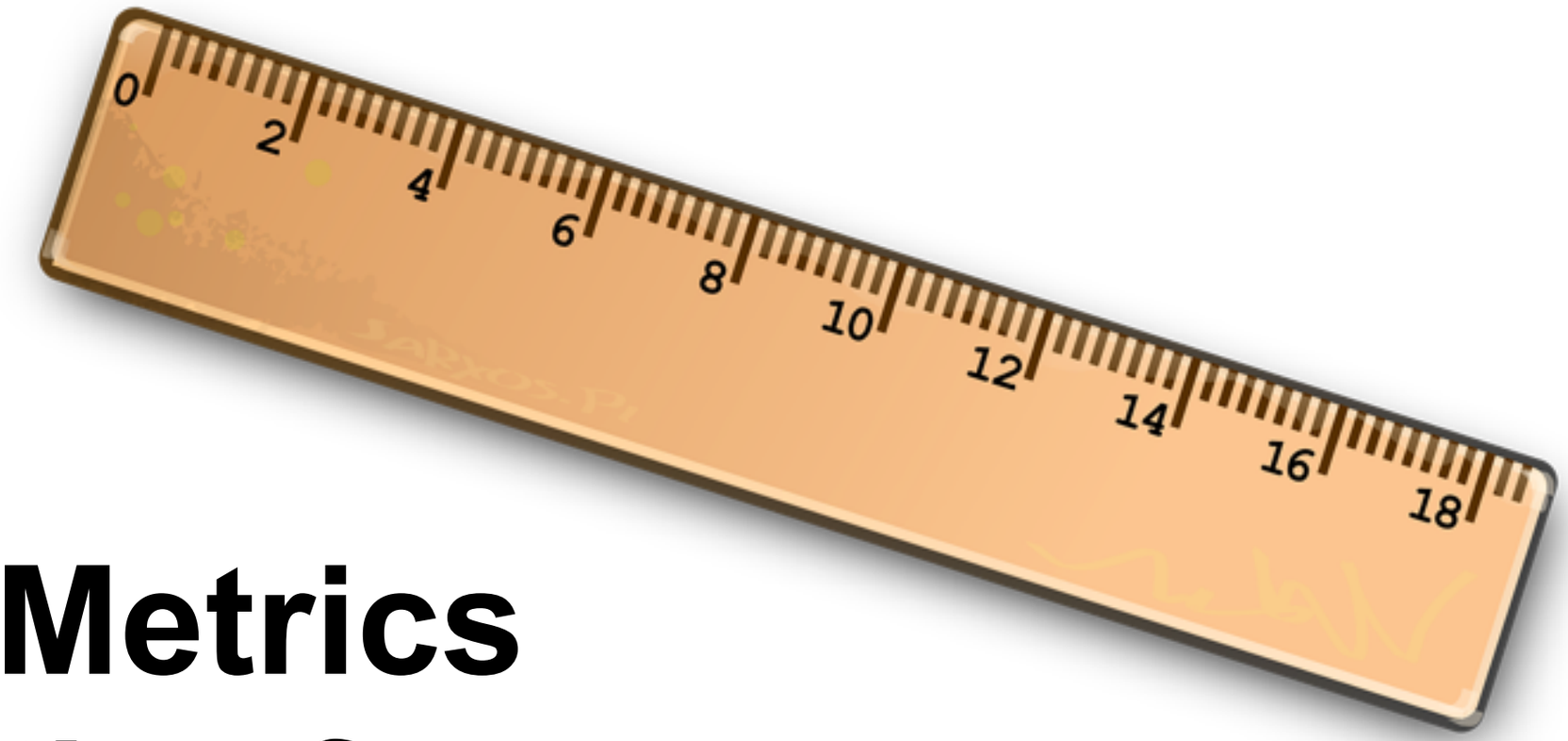


Journal Metrics Perspective from an Open Access Publisher

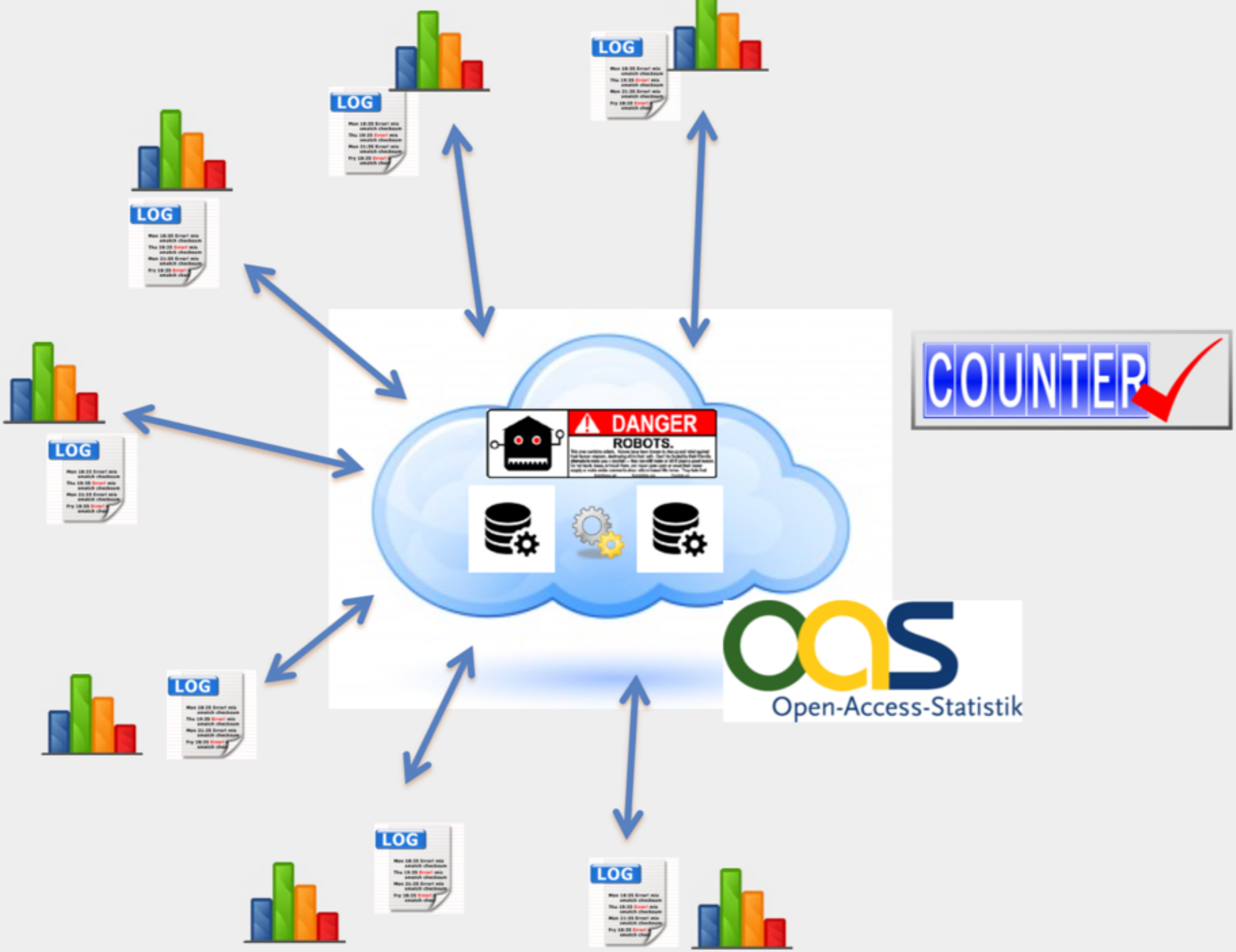
Martin Fenner
Technical Lead Article-Level Metrics
Public Library of Science

http://pixabay.com/en/ruler-straight-edge-tool-geometry-145940/

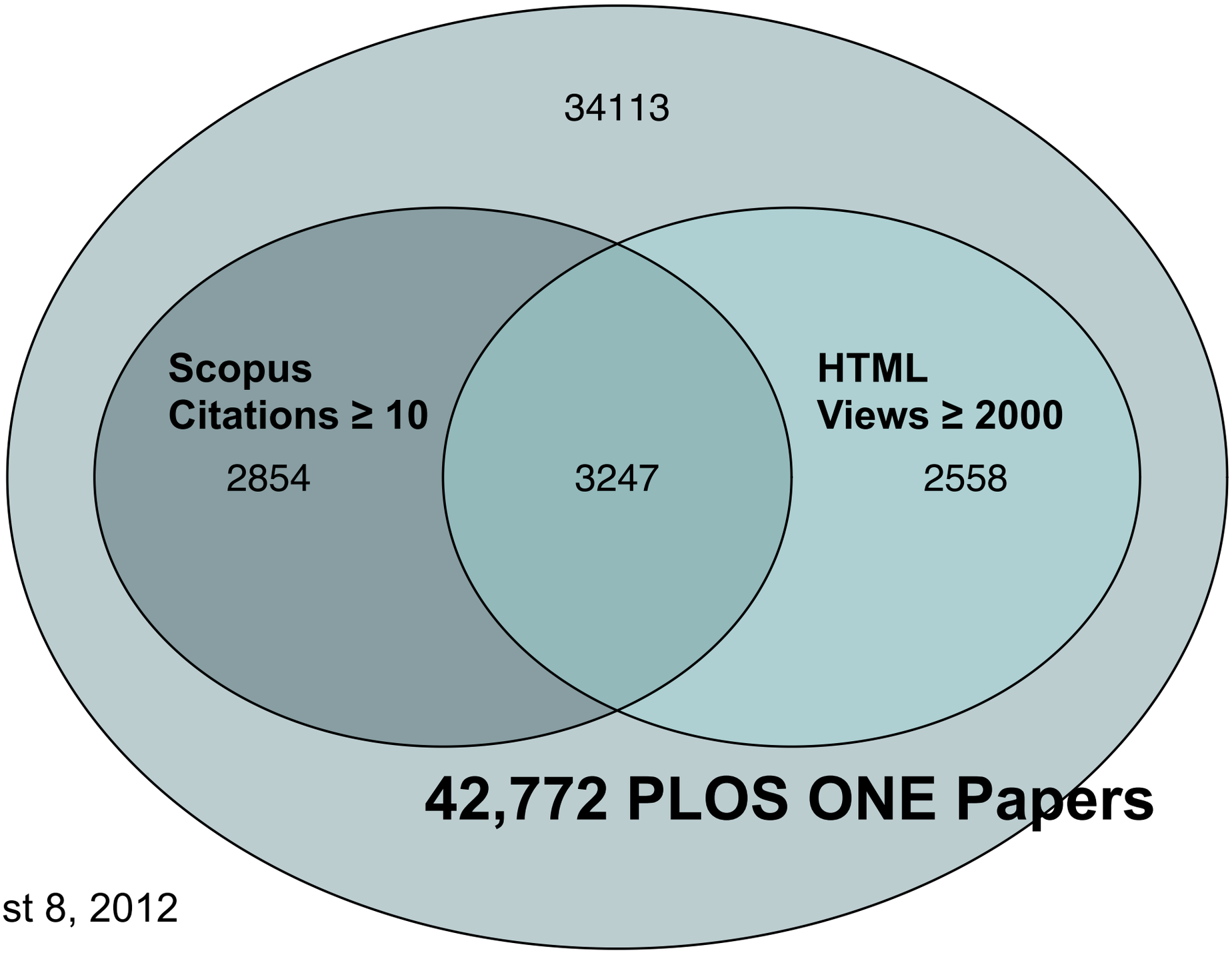
Usage Stats

Most immediate metric that directly reflects usage

Only useful if data are collected in a standardized way – COUNTER is the standard and looks at HTTP status codes, double-click intervals, and excludes robots


LOG
DANGER
ROBOTS.
COUNTER
oas
Open-Access-Statistik

Usage is different from scholarly citations



Metrics collected August 8, 2012

Citations

Citations have become a proxy for scholarly impact

Many problems with unreflected use of citation metrics,
in particular in the assessment of individual researchers

Citations are collected via reference lists

References

1. Kahn P, Hazledine D (1988) NAR's new requirement for data submission to the EMBL data library: information for authors. Nucleic Acids Res 16(10): I–IV.
2. Science as an Open Enterprise (2012) The Royal Society. Available: http://royalsociety.org/policy/projects/science-public-enterprise/report/. Accessed 2013 Apr 8.
3. Rebholz-Schuhmann D, Arregui M, Gaudan S, Kirsch H, Yepes AJ (2007) Text processing through Web services: calling Whatizit. Bioinformatics 24(2):296–298.
4. McEntyre JR, Ananiadou S, Andrews S, Black WJ, Boulderstone R, et al. (2011) UKPMC: a full text article resource for the life sciences. Nucleic Acids Res 39:d58–65.
5. Névéol A, Wilbur WJ, Lu Z (2011) Extraction of data deposition statements from the literature: a method for automatically tracking research results. Bioinformatics 27(23):3306–3312.
6. Névéol A, Wilbur WJ, Lu Z (2012) Improving links between literature and biological data with text mining: a case study with GEO, PDB and MEDLINE. Database (Oxford) 2012: bas026.
7. Fink JL, Kushch S, Williams PR, Bourne PE (2008) BioLit: integrating biological literature with databases. Nucleic Acids Res 36(Web Server Issue):W385–9.
8. Haeussler M, Gerner M, Bergman CM (2011) Annotating genes and genomes with DNA sequences extracted from biomedical articles. Bioinformatics 27(7):980–6.
9. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res Database Issue 38:D211–222.
10. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, et al. (2005) ArrayExpress – a public repository for microarray gene expression data at the EBI. Nucleic Acids Res (2005) 33 (Suppl 1): D553–D555.
11. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. Brief Bioinform (3):225–35.

Reference lists have to be collected in a central resource and in a standard format

CrossRef's specific mandate is to be the citation linking backbone for all scholarly information in electronic form

Limitations

CrossRef citation linking built around references that have DOIs from CrossRef members – usually scholarly articles

CrossRef *Cited-By* service only available to CrossRef members for their own articles

CrossRef is a non-profit organization with publishers as members – no academic institutions, funders, other stakeholders

Alternative citation indexes for non-publisher users

THOMSON REUTERS
WEB OF KNOWLEDGE℠

Scopus

Google scholar

Europe PubMed Central

SciELO
Scientific Electronic Library Online

Reference lists increasingly contain non-article references

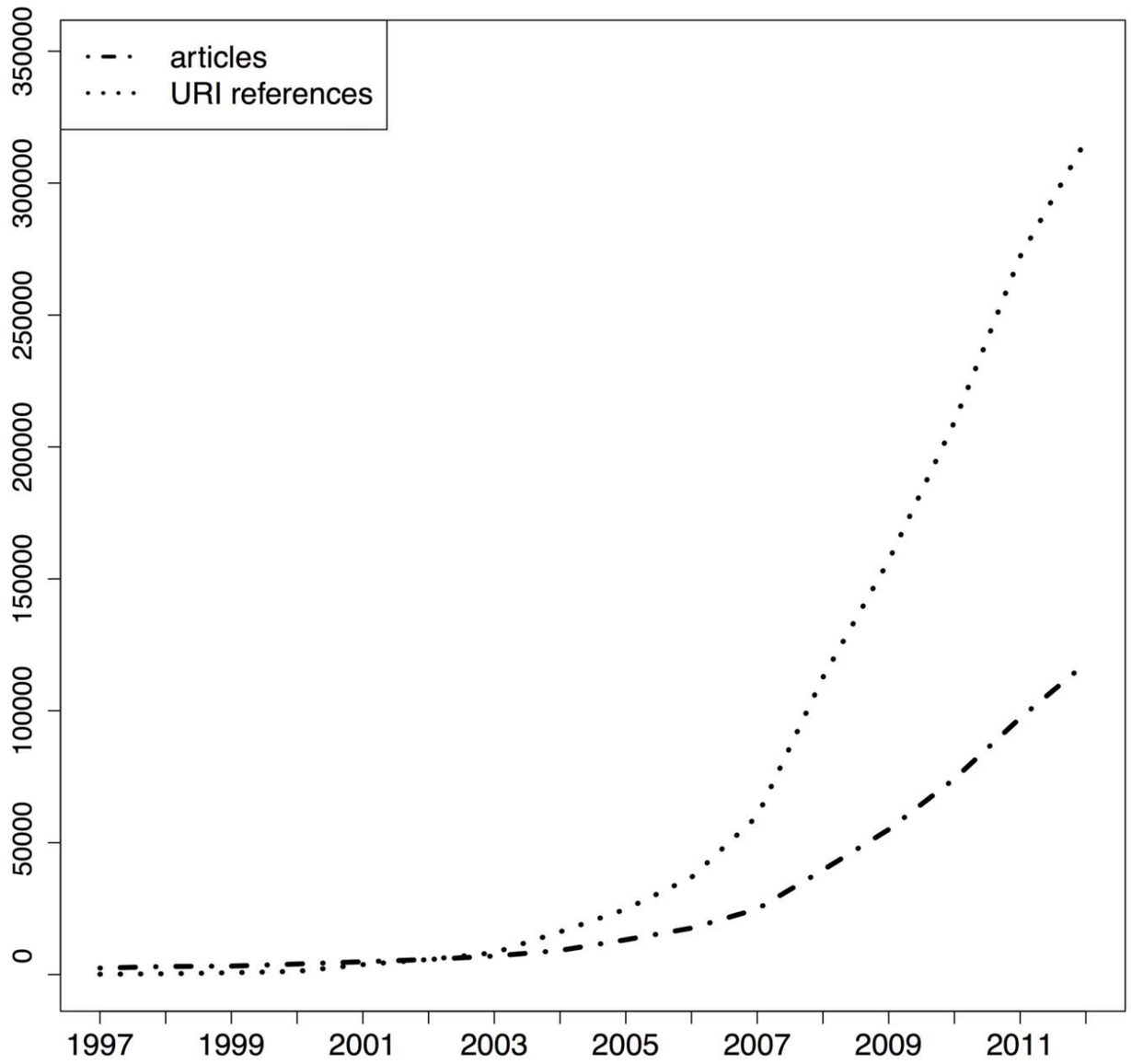


Fig. 3. STM articles and URI references per publication year - PMC corpus.

doi:10.1371/journal.pone.0115253.g003

DataCite provides DOIs for academic institutions and data centers

DataCite

Helping you to find, access, and reuse data

DataCite DOIs are not just for datasets

Total number of indexed DOIs	4,093,327
Number of indexed datasets	901,920
Number of indexed text documents	289,067
Number of indexed software	4,341

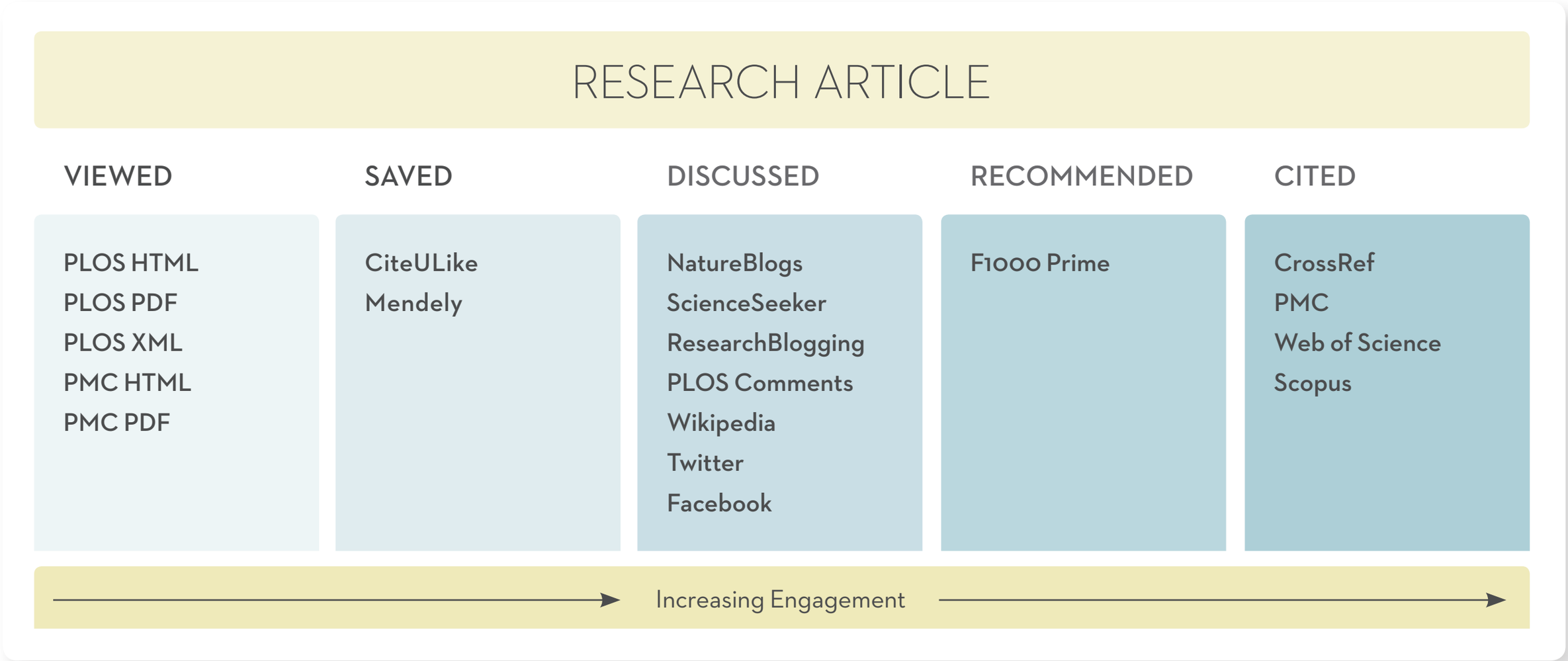
DataCite citation linking works differently, and is separate from CrossRef

A DOI is not a DOI

CrossRef and DataCite announce initiative Nov 2014

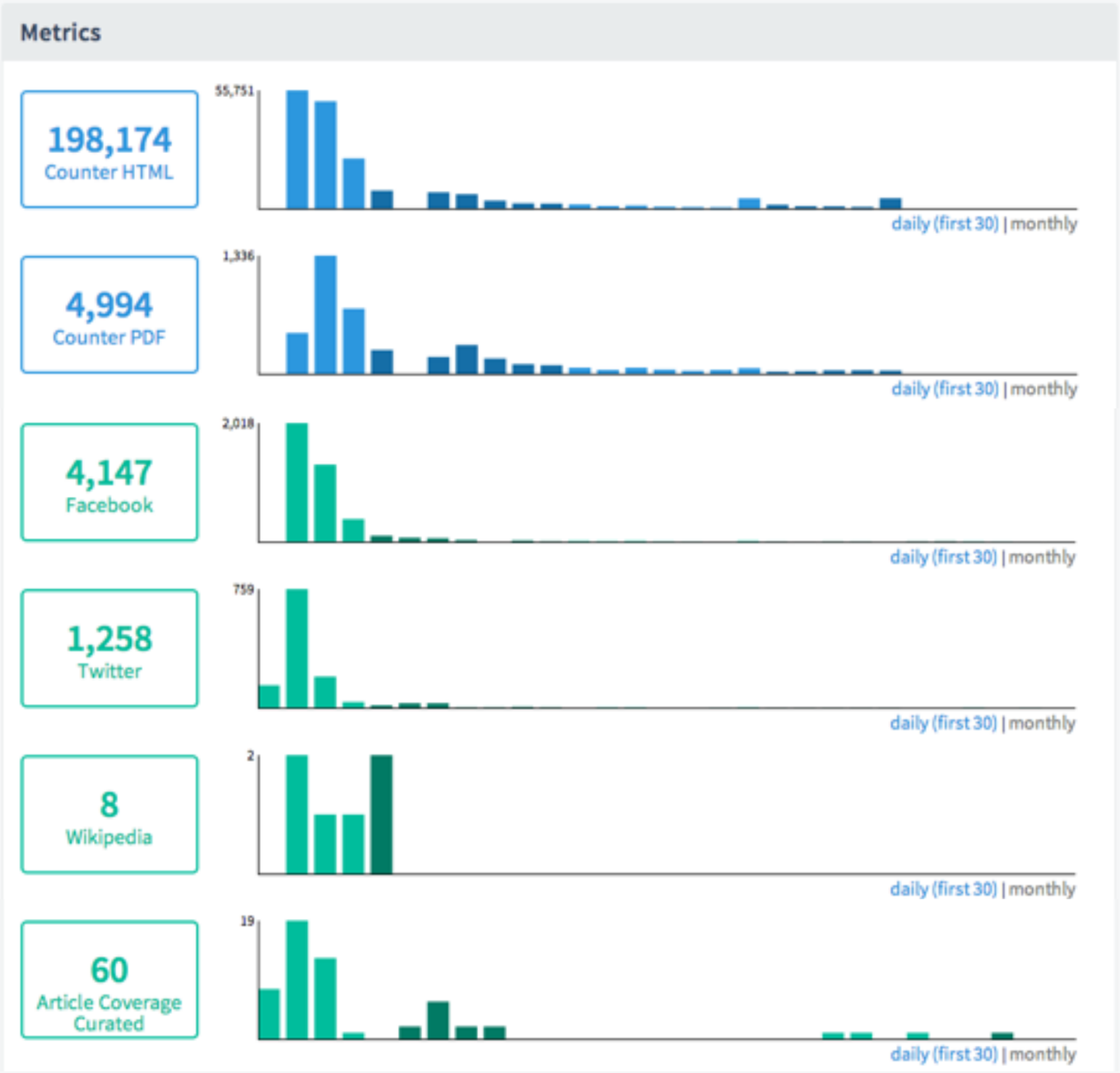
- *Provide comprehensive support for interlinking between articles and data.*
- *Develop open APIs and open source tools to surface citations and other relationships between publications and data sets.*
- *Integrate into their services other existing scholarly communications initiatives such as ORCID and FundRef.*
- *Develop systems, workflows and best practices for using DOIs to reference large, highly granular and dynamic data.*

There is more than usage stats and citations



http://dx.doi.org/10.3789/isqv25no2.2013.04

PLOS collects metrics from 22 data sources



Days since publication
August 27, 2014

New metrics not ready for impact assessment

Unreliable & Unvalid

Unreliable, But Valid

Reliable, Not Valid

Both Reliable & Valid

Any metric we use should have good **reliability** (consistency) and **validity**.

More work is needed in these areas for novel assessment metrics such as Mendeley or Twitter.

http://en.wikipedia.org/wiki/Validity_(statistics)#mediaviewer/File:Reliability_and_validity.svg

Work on best practices and standards has started

NISO
How the information world CONNECTS

Alternative Metrics Initiative
Phase 1
White Paper

June 6, 2014

Phase II of the project starts in early 2015

Altmetrics data can be obtained from commercial service providers

Altmetric

PLUM ANALYTICS

Open source software to collect and analyze the metrics

Lagotto

build failing | code climate 2.0 | coverage 79%

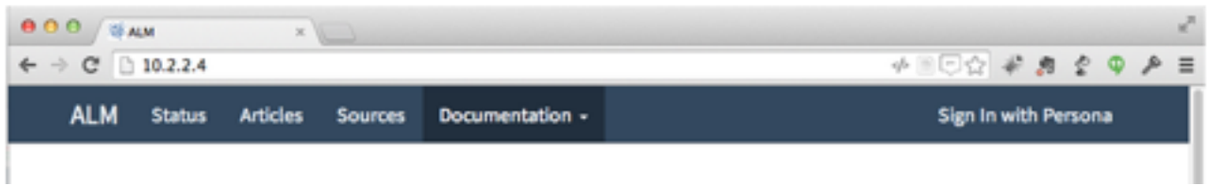
Lagotto allows a user to aggregate relevant performance data on research articles, including how often an article has been viewed, cited, saved, discussed and recommended. The application was called Article-Level Metrics (ALM) until September 2014 and was started in March 2009 by the Open Access publisher Public Library of Science (PLOS). We are continuing to expand Lagotto because we believe that articles should be considered on their own merits, and that the impact of an individual article should not be determined by the journal in which it happened to be published. As a result, we hope that new ways of measuring and evaluating research quality (or 'impact') can and will evolve. To learn more about Article-Level Metrics, see the SPARC ALM primer.

How to start developing now?

`Lagotto` uses Vagrant and Virtualbox for setting up the development environment. To start developing now on your local machine (Mac OS X, Linux or Windows):

1. Install Vagrant: https://www.vagrantup.com/downloads.html
2. Install Virtualbox: https://www.virtualbox.org/wiki/Downloads
3. Clone this repository `git clone git@github.com:articlemetrics/lagotto.git`
4. Cd into it
5. Copy the file `.env.example` to `.env` and make any changes to the configuration as needed
6. Run `vagrant up`

Once the setup is complete (it might take up to 15 minutes), you'll be able to open up a browser and navigate to http://10.2.2.4, and you should see this screen:



https://github.com/articlemetrics/lagotto

ALM Reports

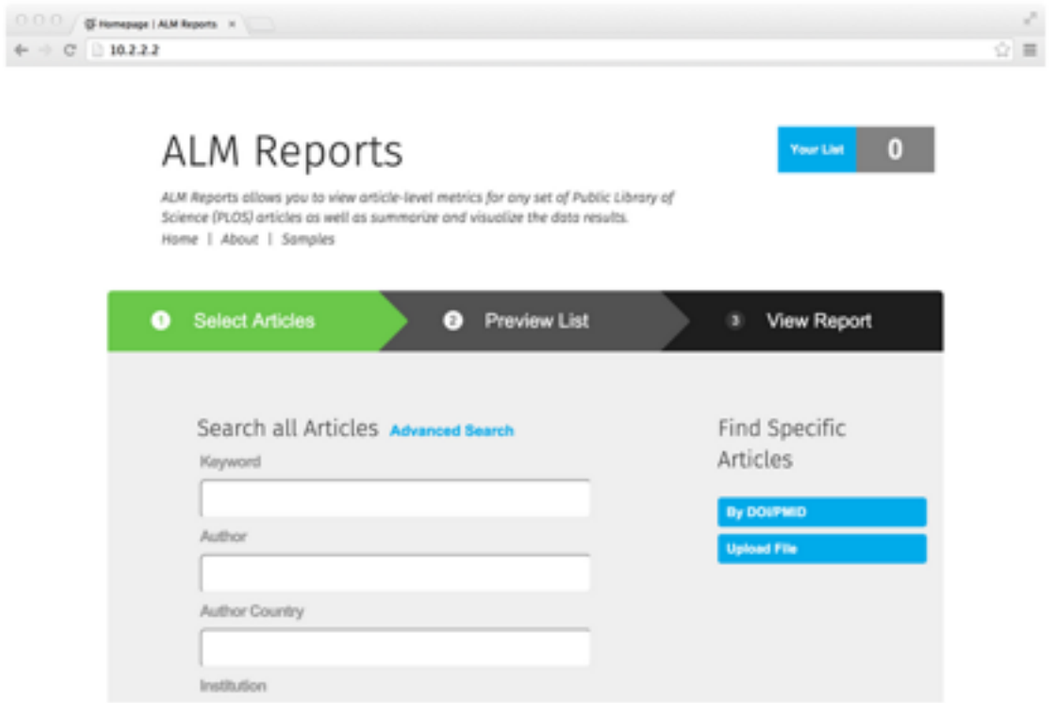
build failing | code climate 1.7 | coverage 73%

How to start developing now?

`ALM Reports` uses Vagrant and Virtualbox for setting up the development environment. To start developing now on your local machine (Mac OS X, Linux or Windows):

1. Install Vagrant: https://www.vagrantup.com/downloads.html
2. Install Virtualbox: https://www.virtualbox.org/wiki/Downloads
3. Clone this repository `git clone git@github.com:articlemetrics/alm-report.git`
4. Cd into it and run `vagrant up`

Once the setup is complete (it might take up to 20 minutes), you'll be able to open up a browser and navigate to http://10.2.2.2, and you should see this screen:



https://github.com/articlemetrics/alm-report

PLOS

CrossRef DOI Event Tracker (DET) Pilot

CrossRef Labs has started a pilot project to collect events around all CrossRef DOIs issued since January 2011.

A DOI Event Tracker (DET) CrossRef working group was formed in May 2014, initiated by members of the Open Access Scholarly Publishers Association (OASPA).

The service is available at http://det.labs.crossref.org, and is using the Lagotto open source software.