

Quality challenges of altmetric data aggregators

Zohreh Zahedi

Centre for Science and Technology Studies (CWTS),
Leiden University, the Netherlands



Universiteit
Leiden

**COAR *metrics Workshop, 14-15 May 2018,
ZBW, Hamburg, Germany.**

Outline

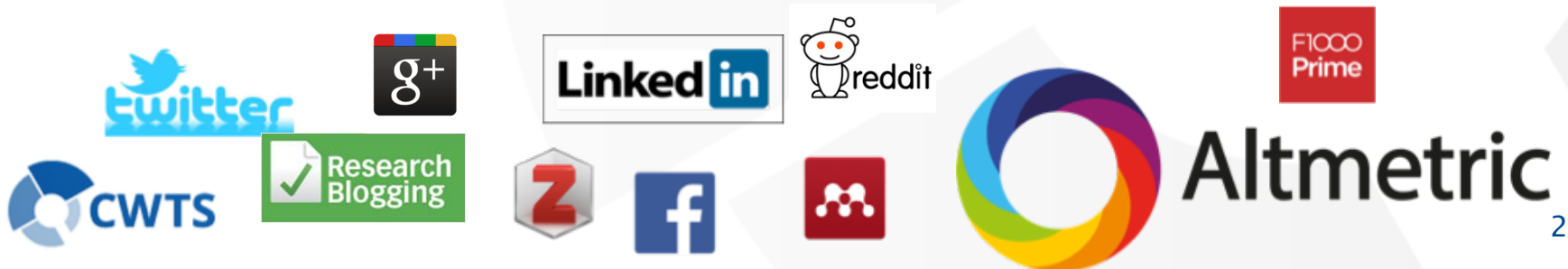
Altmetrics (*social media metrics*) main challenges:

- Definition and concept: what do they mean?
- Type of indicators: what do they indicate?
- **Data quality issues: what are their data quality?**

Altmetrics: main ideas and concept(s)

Altmetrics Manifesto (2010) (<http://altmetrics.org/manifesto/>)

- **Several terms:** alternative metrics, altmetrics, article level metrics, influmetrics, social media metrics (Haustein et al., 2015; Wouters, Zahedi, & Costas, 2018)
- **Several definitions** (general and specific and no consensus):
 - ‘the creation and study of new metrics based on the Social Web for tracking, analyzing, and informing scholarship or as a form of information filtering tool’ (Priem et al., 2010).
 - online events derived from activity and engagement between diverse stakeholders and scholarly outputs in the research ecosystem (NISO, 2016).
 - events on social and mainstream media platforms related to scholarly content or scholars [...] and are not the same as [...] citations (Haustein, Bowman, Costas, 2016).
- **Heterogeneous acts:** refers to diverse activities in different platforms: blogs, Twitter, Facebook, Wikipedia, reference management tools, etc.



Some attempts to classify or characterize social media metrics

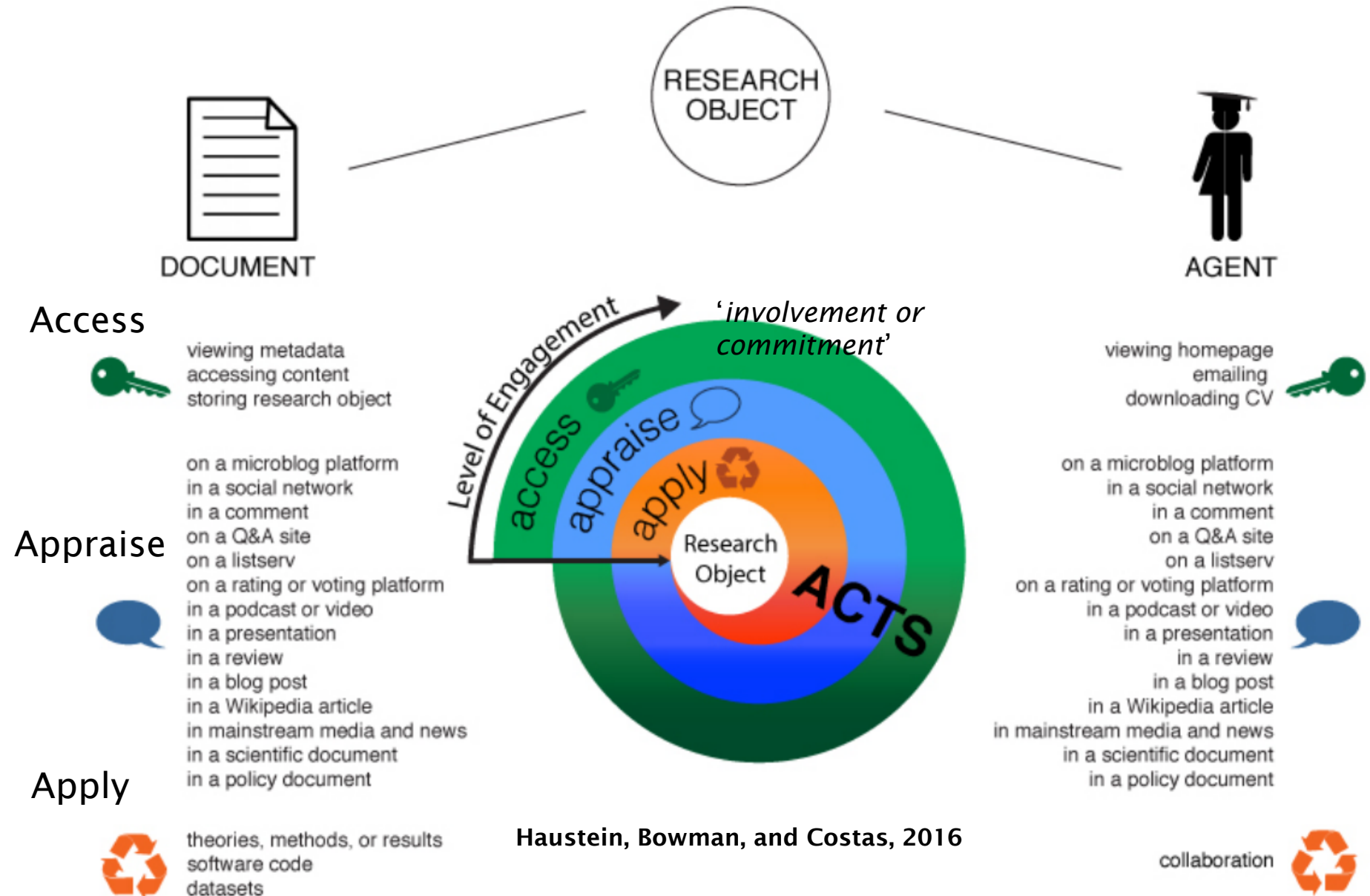
- Frameworks of acts around research objects (Haustein, Bowman, and Costas, 2016).

Haustein, S., Bowman, T. D., & Costas, R. (2016). *Interpreting Altmetrics: Viewing Acts on Social Media through the Lens of Citation and Social Theories*. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication* (pp. 372–406). Berlin, Boston: De Gruyter. <http://doi.org/10.1515/9783110308464-022>

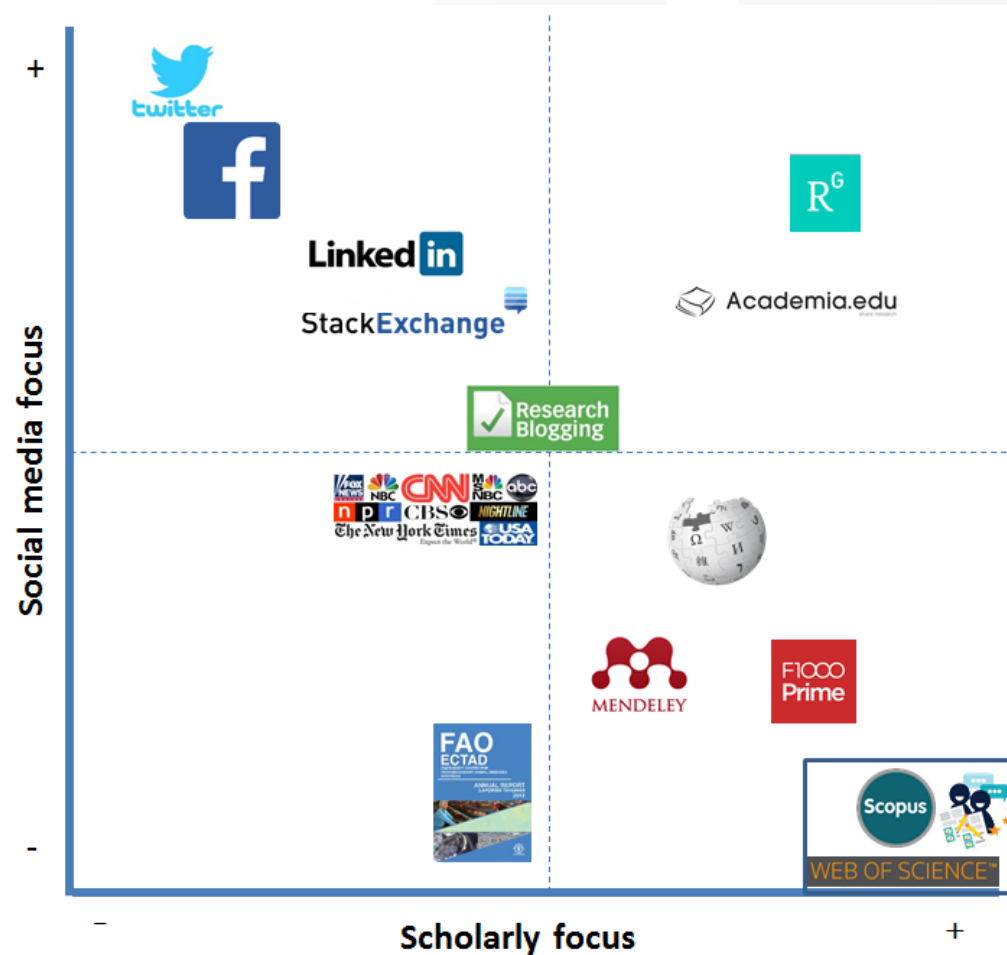
- Characterizing metrics by their social media or scholarly *focus* (Wouters, Zahedi, & Costas, 2018).

Wouters, P., Zahedi, Z., & Costas, R. (2018). *Social media metrics for new research evaluation*. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of Quantitative Science and Technology Research*. Springer.

Framework of 'acts' around Research Objects



Characterizing metrics by their social media or scholarly focus



Other challenges of social media metrics:

- Type of indicators: what do they indicate?
 - Diverse and heterogeneous: What type of impact they refer? As an indicator of (early) scientific impact, societal impact, social recognition, online attention, etc.
 - Fast: Is faster always better? Superficiality? what type of engagement they refer?
- **Data quality issues: what are their data quality?**
 - Openness: collection of metrics through open API. How transparent and consistent are they?
 - **How are data quality of altmetric data?**

Considerations of altmetrics for research evaluation

Altmetrics offer the possibility of studying new forms of *interactions* between *social media users*, *scholarly objects*, and different *academic actors* (Costas, 2017).

- Indicators need “to adhere to a far stricter protocol of data quality and indicator *reliability and validity*” (Wouters & Costas, 2012).

Reliability is the extent to which an indicator yields the same result in repeated measurements.

Validity relates to an indicator's success at measuring what is expected to be measured.

- *To what extent social media indicators are valid and reliable as measures of research performance?*
- *How is the data quality of altmetric data provided by different aggregators?*

Indicators should meet three essential criteria to be *valid: adequacy, sensitivity, and homogeneity* (Gingras, 2014).

- *Adequacy* indicates how the indicator captures the reality behind the concept intended to be measured.
 - it is not clear to what extent social media indicators are related to the concept of scientific impact and valid as measures of research performance.
- Sensitivity *relates to the resistance to change of indicators*.
 - It is challenged by their speed, superficiality, and reliability.
 - For example, indicators based on social media platforms like Twitter can drastically change in a matter of hours while bibliometric indicators don't suffer from sudden changes.
- Homogeneity is important with respect to composite indicators.
 - Altmetric Attention Score or the RG score which lump together fundamentally different metrics (e.g. Twitter, blogs, views, etc.) .

Recommendations for altmetric data aggregators: *Transparency, replicability, and accuracy* (NISO, 2016).

- *Transparency* refers to the degree to which information and details about the provided data are clear, well-documented, and open to all users (human and machine) for verification.
- *Replicability* refers to the degree to which a set of data is consistent across providers and aggregators and over time.
- *Accuracy* refers to the degree to which the collected data reflects the material it claims to describe.

NISO code of conduct for altmetrics

Altmetric data providers are encouraged, and altmetric data aggregators are expected to offer:

Transparent data by offering information about:

- how data are generated, collected, and curated
- how data are aggregated, and derived data generated

Replicable data by ensuring that:

- the provided data is generated using the same methods over time;
- changes in methods and their effects are documented;
- changes in the data following corrections of errors are documented;
- data provided to different users at the same time is identical or, if not, differences in access provided to different user groups are documented.

Accurate data by ensuring that:

- the data represents what it supposed to reflect.
- known errors are identified and corrected any limitations of the provided data are communicated.

Inconsistency among altmetric data aggregators

Studies have showed that how the social media metrics collected for a same set of DOIs at the same time may vary across different major altmetric data aggregators (Meschede & Siebenlist, 2018; Zahedi & Costas, 2018; Ortega, 2017; Bar-Ilan and Halevi, 2017, Chamberlain, 2013; Zahedi, Fenner, and Costas, 2014; 2015; Jobmann, et al., 2014).

- For instance there are differences in the coverage, total counts, and correlation of metrics across different aggregators.

Reasons for the difference across altmetric data aggregators

- Data collection choices (use of different APIs, different identifiers, URLs, landing pages, or scholarly objects);
- **Data aggregation and reporting choices** (based on different identifiers, versions, languages, edits, etc.);
- Updating choices (possible time lags in the data collection or updating frequency);
- Other technical choices (accessibility issues such as cookies problems or access denies for resolving DOIs to URLs differ across different publisher platforms).

(Zahedi & Costas, 2018; Zahedi, Fenner, and Costas, 2014; 2015)

Social media metrics can largely be influenced by

- Aggregations of different identifiers (DOI, PMID, ArXiv ID, URL) and versions (e.g. the ArXiv versions, the published version) of the same publication.
- Aggregations of different events from the same social media platform (Facebook likes, shares, comments or tweets and retweets).
- Aggregation based on languages, document types, scholarly objects, and edits.

Inconsistency in the readership counts across different aggregators

Even using the same API the specific way each aggregator queries the *API using specific or multiple identifiers* can have an effect on the final reported counts.

- For instance, Altmetric.com only queries Mendeley readership when other altmetric event has been reported for the publication.
- Altmetric.com queries Mendeley first using the DOI of the publication, and if no record is found then the PMID, and subsequently the ArXiv id is used. It stops the API when one of the identifiers has provided a match.
- Lagotto queries Mendeley API using only DOIs.
- Plum Analytics does not query Mendeley API since it gets the data from Mendeley as it is part of Elsevier.

- Plum Analytics merges readership counts from different identifiers (e.g. DOI and PMIDs) for the same publication.
- when the same object appears with different records in Mendeley, for example one with a DOI and another one with a PMID, Plum Analytics aggregates all the readership counts across all the different versions.
- The problem is that the merging of counts from different identifiers can also imply some degree of error. For example, wrong linkages between identifiers happen and may create over or under-merging of records.

An example of a record in Mendeley with correct DOI:



The screenshot shows a Mendeley record for a journal article. The Mendeley logo and navigation links are at the top. The article title is "Marine communities on oil platforms in Gabon, West Africa: High biodiversity oases in a low biodiversity environment". The authors are Friedlander A, Ballesteros E, Fay M, Sala E. The journal is PLoS ONE, vol. 9, issue 8 (2014). The article has 89 Readers and 16 Citations, with the Readers count highlighted by a red box. Below the article information, there are tabs for "Overview" and "Authors (4)". The "Abstract" tab is selected, showing a paragraph of text. To the right of the abstract, there is a section titled "Find this document" which lists various identifiers: DOI, ISSN, SGR, SCOPUS, arXiv, PUI, PMID, and ISBN.

Mendeley What is Mendeley? Search Create a free account Sign In

JOURNAL ARTICLE

Marine communities on oil platforms in Gabon, West Africa: High biodiversity oases in a low biodiversity environment

89 Readers 16 Citations

Friedlander A, Ballesteros E, Fay M, Sala E
PLoS ONE, vol. 9, issue 8 (2014) Published by Public Library of Science

+ Sign in to save reference


Overview Authors (4)

Abstract


The marine biodiversity of Gabon, West Africa has not been well studied and is largely unknown. Our examination of marine communities associated with oil platforms in Gabon is the first scientific investigation of these structures and highlights the unique ecosystems associated with them. A number of species previously unknown to Gabonese waters were recorded during our surveys on these platforms. Clear distinctions in benthic communities were observed between older, larger platforms in the north and newer platforms to the south or closer to shore. The former were dominated by a solitary cup coral, *Tubastraea* sp., whereas the latter were dominated by the barnacle *Megabalanus tintinnabulum*, but with more diverse benthic assemblages compared to the northerly platforms. Previous work documented the presence of limited zooxanthellated scleractinian corals on natural rocky substrate in Gabon but none were recorded on platforms. Total estimated fish biomass on these platforms

Find this document

DOI: 10.1371/journal.pone.0103709
ISSN: 19326203
SGR: 84905369106
SCOPUS: 2-s2.0-84905369106
arXiv: arXiv:1011.1669v3
PUI: 373686004
PMID: 25083704
ISBN: 1576 – 9526(colección); 84 – 9224...

 <https://www.mendeley.com/research-papers/marine-communities-oil-platforms-gabon-west-africa-high-biodiversity-oases-low-biodiversity-envirom/>

Plum Analytics aggregates readership (211 reader counts instead of 89 reader counts from Mendeley) for the paper with both correct and incorrect DOIs and PMIDs in Mendeley



Marine communities on oil platforms in Gabon, West Africa: high biodiversity oases in a low biodiversity environment.

Citation data: PloS one, ISSN: 1932-6203, Vol: 9, Issue: 8, Page: e103709
Publication Year: 2014

Explore PlumX Metrics

What are PlumX Metrics? How can they help tell the story about this research? How can I use them?

[Learn more](#)

USAGE ▾	5793	CAPTURES ▾	252	SOCIAL MEDIA ▾	36	CITATIONS ▾	16
Full Text Views ⓘ	4999	Readers ⓘ	211	Shares, Likes & Comments ⓘ	32	Citation Indexes ⓘ	16
Abstract Views ⓘ	573	Mendeley	116	Tweets ⓘ	4		
Views ⓘ	190	Mendeley	89				
Downloads ⓘ	23	Mendeley	3				
Link-outs ⓘ	8	Mendeley	3				
		Exports-Saves ⓘ	41				

ARTICLE SUMMARY

TWEETS



PMID:

25083704

DOI:

10.1371/journal.pone.0103709.t006, 10.1371/journal.pone.0103709.t007,

MOST RECENT TWEET

  **melissa goulton**
@MelGoulton

[View All Tweets](#)

An example of a record in Mendeley with incorrect PMDI:

Secure | <https://www.mendeley.com/research-papers/métodos-para-medir-la-biodiversidad-23/>

etric it | Locations Leiden Uni | Save to Mendeley | FSW webmail | Google Maps | Google Calendar | UBL Get It | Google Translate | Google Scholar | Dashboard | Mendeley | UBL Get It | Re

Mendeley What is Mendeley? Q Search Create a free account Sign In

JOURNAL ARTICLE

Métodos para medir la biodiversidad

116 Readers N/A Citations

Moreno C
M&T Manuales y Tesis Sociedad Entomológica Aragonesa, vol.1., vol. 1 (2001) p. 84

+ Sign in to save reference

Overview Authors (1)

Abstract

Todos los sistemas biológicos son diversos. Es decir, varían en el número y cantidad de las partes que los forman. La diversidad biológica o biodiversidad es la propiedad de la vida, a distintos niveles de organización, de ser diversa. Así, los individuos de una especie muestran diferencias en la estructura de su ácido desoxirribonucleico (ADN), la molécula que codifica la información genética. A esta variabilidad se le conoce como diversidad genética. En otro nivel de organización, las comunidades ecológicas están integradas por un determinado número de especies, y cada una de estas especies tiene una cierta importancia en la comunidad. Dicha importancia está determinada por el número de individuos, biomasa, cobertura, etc. de cada una de las especies. A esta variabilidad se le conoce como diversidad de especies. Finalmente, las áreas geográficas que se distinguen como "paisajes" por tener una historia y condiciones ambientales particulares, están integradas por distintos hábitats que intercambian materiales bióticos y abióticos. El número y representatividad de estos hábitats constituyen la diversidad de ecosistemas en el paisaje. Durante las últimas décadas han surgido numerosas propuestas de métodos para medir las diferentes facetas de la biodiversidad. Pero a pesar de que la

Find this document

arXiv: arXiv:1011.1669v3
ISSN: 19326203
ISBN: 84-927495-2-8
PMID: 25083704
DOI: 10.1371/journal.pone.0103709

Get full text

Authors

CM Claudia Elizabeth Moreno

- The higher readership values reported by Lagotto in comparison to the other aggregators can be explained by Lagotto's choice of reporting combined readership values of individual user and group counts.



Minimum pricing of alcohol versus volumetric taxation: which policy will reduce heavy consumption without adversely affecting light and moderate consumers?

Citation data: PLoS one, ISSN: 1932-6203, Vol: 9, Issue: 1, Page: e80936
Publication Year: 2014

USAGE ▾	8540	CAPTURES ▾	70	MENTIONS ▾	1	SOCIAL MEDIA ▾	92	CITATIONS ▾	19
HTML Views ▾	7401	Exports-Saves ▾	39	News Mentions ▾	1	Tweets ▾	92	Citation Indexes ▾	19
PDF Views ▾	836	Readers ▾	31						
Abstract Views ▾	246	Mendeley	31						
Views ▾	30								
Link-outs ▾	16								

Explore PlumX Metrics

What are PlumX Metrics? How can they help tell the story about this research? How can I use them?

[Learn more](#)

ALM Status Works Sources Support ▾ Sign in with PLOS ID

Minimum Pricing of Alcohol versus Volumetric Taxation: Which Policy Will Reduce Heavy Consumption without Adversely Affecting Light and Moderate Consumers?

[View All Tweets](#)

Publication Date	January 22, 2014
Journal	PLOS ONE
Authors	Anurag Sharma, Brian Vandenberg & Bruce Hollingsworth
Volume	9
Issue	1
Pages	e80936
DOI	http://doi.org/10.1371/journal.pone.0080936
Publisher URL	http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0080936
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/24465368
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898955
Europe PMC	http://europepmc.org/abstract/MED/24465368
Web of Science	000330283100002
Scopus	84899858219
Mendeley	http://www.mendeley.com/research/minimum-pricing-alcohol-consumption-without-adversely-affecting-light-and-moderate-consumers/

References Events Details Recommendations

CiteULike

Mendeley 34

Minimum Pricing of Alcohol versus Volumetric Taxation: Which Policy Will Reduce Heavy Consumption without Adversely Affecting Light and Moderate Consumers?

Overview of attention for article published in PLoS ONE, January 2014



About this Attention Score

In the top 5% of all research outputs scored by Altmetric

MORE...

Mentioned by

- 2 news outlets
- 1 blog
- 1 policy source
- 72 tweeters
- 2 Facebook pages

Readers on

- 31 Mendeley

SUMMARY

News

Blogs

Policy documents

Twitter

Facebook

You are seeing a free-to-access but limited selection of the activity Altmetric has collected about this research output. [Click here to find out more.](#)

Title	Minimum Pricing of Alcohol versus Volumetric Taxation: Which Policy Will Reduce Heavy Consumption without Adversely Affecting Light and Moderate Consumers?
Published in	PLoS ONE, January 2014
DOI	10.1371/journal.pone.0080936
Pubmed ID	24465368
Authors	Anurag Sharma, Brian Vandenberg, Bruce Hollingsworth, Sharma A, Vandenberg B, Hollingsworth B
Abstract	We estimate the effect on light, moderate and heavy consumers of alcohol from implementing a... [show]

[View on publisher site](#)

[Alert me about new mentions](#)

TWITTER DEMOGRAPHICS

MENDELEY READERS

ATTENTION SCORE IN CONTEXT

The data shown below were collected from the profiles of 72 tweeters who shared this research output. [Click here to find out more about how the information was compiled.](#)



Inconsistency in the Facebook counts across different aggregators

Altmetric.com, Lagotto, and Plum Analytics use the same Facebook Graph API.

- the choice of aggregators in collecting public or private scores also results in variations in the metrics offered by them.
- For instance, Lagotto and Plum Analytics collect both Facebook public and private posts while Altmetric.com collects only public wall posts.
- Additionally Plum Analytics and Lagotto also count other Facebook events such as likes, shares, and comments.



Nutrition and health - the association between eating behavior and various health parameters: a matched sample study.

Citation data: PLoS one, ISSN: 1932-6203, Vol: 9, Issue: 2, Page: e88278
Publication Year: 2014

Explore PlumX Metrics

What are PlumX Metrics? How can they help tell the story about this research? How can I use them?

[Learn more](#)

USAGE ▼

526745

HTML Views	298676
PDF Views	224370
Abstract Views	2769
Clicks	788
Views	89
Link-outs	36
Downloads	17

CAPTURES ▼

411

Exports-Saves	408
Readers	3

MENTIONS ▼

2239

Comments	2228
Blog Mentions	7
News Mentions	3
Economics Blog Mentions	1

SOCIAL MEDIA ▲

28690

Shares, Likes & Comments	27296
Facebook	27296

CITATIONS ▼

15

Citation Indexes	15
------------------	----

RATINGS ▼

1592

Reddit	1592
--------	------

ALM Status Works Sources Support ▼

Sign in with PLOS ID

Works

10.1371/journal.pone.008...

[References](#)

Facebook ▼

Sort by Date ▼

Funding sources and conflicts of interest

February 27, 2014. <http://dx.doi.org/10.1371/journal.pone.0088278>
Facebook: NaN

Nutrition and Health – The Association between Eating Behavior and Various Health Parameters: A Matched Sample Study

February 27, 2014. <http://doi.org/10.1371/journal.pone.0088278>
Facebook: 1,023 | Viewed: 524,107 • Saved: 268 • Discussed: 1,955

SUMMARY

News

Blogs

Twitter

Web

So far, Altmeter has seen **264** public wall posts from **261** users.

Showing items 1–100

1 2 3



Michal Piják, MD - Personalised Evolutionary Nutrition, 14 Sep 2017
VÝŠŠÍ VÝSKYT RAKOVINY, ALERGICKÝCH A MENTÁLNÝCH OCHORENÍ O VEGETARIÁNŮV



SCPHRP - Scottish Collaboration for Public Health Research and Policy, 28 Apr 2017
Research: Nutrition and Health – The Association between Eating Behavior and Various Health Parameters <https://t.co/hjmM5uyZhm>

Altmeter

Nutrition and Health – The Association between Health Parameters: A Matched Sample Study

Overview of attention for article published in PLoS ONE, February 2014

971

About this Attention Score

In consistency in the tweet counts across different aggregators

Lagotto reports the combined tweets and retweets.

Plum Analytics also reports combined scores of tweets and retweets for the different versions of the same object.

CrossRef ED also reports all data of tweets and retweets.

Altmetric.com reports the number of distinct tweeters.



Nutrition and health - the association between eating behavior and various health parameters: a matched sample study.

Citation data: PLoS one, ISSN: 1932-6203, Vol: 9, Issue: 2, Page: e88278
Publication Year: 2014

Explore PlumX Metrics

What are PlumX Metrics? How can they help tell the story about this research? How can I use them?

Learn more

USAGE ▾	540279	CAPTURES ▾	411	MENTIONS ▾	2239
HTML Views ●	302707	Exports-Saves ●	408	Comments ●	2228
PDF Views ●	233873	Readers ●	3	Blog Mentions ●	7
Abstract Views ●	2769				
Clicks ●	788				
Views ●	89				
Link-outs ●	36				
Downloads ●	17				

SOCIAL MEDIA ▾	28702
Shares, Likes & Comments ●	27296
Tweets ●	1254
Twitter	1254
+1s ●	152

Nutrition and Health – The Association between Eating Behavior and Various Health Parameters: A Matched Sample Study

Overview of attention for article published in PLoS ONE, February 2014



About this Attention Score

In the top 5% of all research outputs scored by Altmetric

Mentioned by

- 4 news outlets
- 9 blogs
- 907 tweeters
- 1 weibo user
- 261 Facebook pages
- 27 Google+ users
- 8 Redditors
- 1 video uploader

Readers on

- 264 Mendeley
- 3 CiteULike

SUMMARY News Blogs Twitter Weibo Facebook Google+ Reddit Video

So far, Altmetric has seen 993 tweets from 907 users, with an upper bound of 2,295,823 followers.

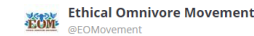
Showing items 1–100



Vegeta
@Vegeta_9000

@theworstkirst @Tomleewalker Further there is conflicting evidence about the health risk of vegan diet. <https://t.co/mjuqEX0rFV>

25 Nov 2017



Ethical Omnivore Movement
@EOMovement

@robbwolf for the win <https://t.co/oTxIOAgPJ>

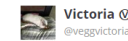
20 Nov 2017



Asnaketch Woldetensa
@AWoldetensa

RT @soil4climate: Thank you @robbwolf. Excellent scholarship. <https://t.co/081fjsLjkl>

19 Nov 2017



Victoria
@veggvictoria

@saulochao306_g <https://t.co/dXNEKe4qX1> This?

25 Nov 2017



Allan Savory
@AllanRSavory

RT @robbwolf: @Sinstruine @EOMovement @soil4climate @AllanRSavory @SavoryInstitute BUT, we now do have studies which have decoupled associa...

20 Nov 2017



Soil4Climate
@soil4climate

Thank you @robbwolf. Excellent scholarship. <https://t.co/081fjsLjkl>

19 Nov 2017

This article has 1252 Twitter interactions

View by URL

View All Tweets

702 | 550

70

Nutrition and Health – The Association between Eating Behavior and Various Health Parameters: A Matched Sample Study

Publication Date	February 07, 2014
Journal	PLOS ONE
Authors	Nathalie T. Burkert, Johanna Muckenhuber, Franziska Großschädl, Éva Rásky, et al
Volume	9
Issue	2
Pages	e88278
DOI	http://doi.org/10.1371/journal.pone.0088278
Publisher URL	http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0088278
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/24516625
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3917888
Europe PMC	http://europepmc.org/abstract/MED/24516625
Web of Science	000330834400049
Scopus	84895769056
Mendeley	http://www.mendeley.com/research/nutrition-health-association-between-eating-behavior-various-health-parameters-

Twitter 941

29 Nov 05:03 UTC



Discrepancies in the value of Wikipedia mentions

- Different approaches in aggregating DOIs mentioned across Wikipedia pages in different *languages* as well as *non-encyclopedia pages* (such as user, talk, and Meta-wiki pages, media and files) and consideration of the *edits* of each Wikipedia page as separate events influences the counts.
- Wikipedia mentions count reported by Lagotto is a combined score of the number of references to papers, material files, images, etc. from the 25 most popular Wikipedia sites (non-encyclopedic pages).
- Altmetric.com reports Wikipedia mentions of scholarly outputs collected from the reference sections of English, and from 2017 onwards Finnish, and Swedish languages Wikipedia entries. Altmetric.com doesn't track non-encyclopedic pages.
- Plum Analytics tracks scholarly objects (thesis, book chapters, books, and technical reports) other than articles as well as the Spanish and Portuguese language Wikipedia entries.
- Wikipedia mentions reported by CrossRef ED includes 'edits of articles in Wikipedia' and combines mentions from all old and new versions, thus every edit of the paper is considered separately.

In summary:

Social media metrics data collection relies on a large range of:

- *different methodological and technical choices* (e.g. APIs, identifiers tracked, forms of querying original sources, types of events recorded, selections of publishers)
- *reporting choices* (e.g. aggregation of different types of counts into one single metric, grouping of different metrics into broader categories, combination of different counts for different identifiers or urls).
- Hence, it is important to understand how these choices may affect the data collected and reported by different aggregators.

Limitations of application of social media metrics in research evolution

- Heterogeneity, data quality, dependencies of social media metrics are major challenges of altmetric data (Haustein, 2016).
- the dependence on unique identifiers of scientific publications (e.g. DOI, PMID, etc.). Publications without any of these identifiers are excluded from the tracking algorithms of altmetric data aggregators.
- the dependence on a direct link to the scientific publications. Mentions of publications using just their titles or other textual characteristics of the publication, as well as links to versions of the publication not covered by the altmetric data aggregators will be ignored.
- the direct dependency on altmetric data aggregators, which themselves are also dependent on other major social media data providers (e.g. Twitter, Facebook, etc.) (Haustein, 2016).
- application of social media metrics is potentially limited by the decisions, strategies and changes of any of altmetric data aggregators as variations in their policies may imply the disappearance of a data source.

Recommendations for altmetric data aggregators

- Increase the transparency around the methodological choices in data collection;
- Increase the transparency around the computation of different social media acts;
- Increase the awareness of unintended effects of methodological choices;
- Increase the replicability and interactivity of the data reported.

Recommendations for researchers, librarians, and policy managers

Be aware of the underlying data:

- How it is collected;
- How it is calculated;
- How it is updated;
- How it is reported.

For more information see:

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: extensive comparison of four major altmetric data aggregators. *PLoS ONE*.

Special thanks to the 'Metrics Project' group for the invitation!

&

Thanks a lot for your attention!



z.zahedi.2@cwts.leidenuniv.nl

@zohrehzahedi