Humboldt-Universität zu Berlin
Institut für Bibliotheks- und Informationswissenschaft

Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft

Heft 444

Altmetrics and Open Access
Comparison of Altmetric Scores of open and closed access articles published by German research institutions in the field of natural sciences

von
Lea Satzinger

Altmetrics and Open Access
Comparison of Altmetric Scores of open and closed access articles published by German research institutions in the field of natural sciences

von
Lea Satzinger

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Vivien Petras
Humboldt-Universität zu Berlin

Heft 444

Satzinger, Lea

Altmetrics and Open Access : Comparison of Altmetric Scores of open and closed access articles published by German research institutions in the field of natural sciences / von Lea Satzinger. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2019. – 81 S. : 15 Abb. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 444)

ISSN 14 38-76 62

Abstract

Altmetrics, in contrast to traditional metrics, measure the societal impact research outputs have on the public in general, using social media platforms as their primary data sources. In this study, differences in Altmetric Scores between open and closed access articles of German research institutions in the field of natural sciences have been analyzed. For this investigation data from the years 2013 to 2017 was gathered from *Web of Science*, *Altmetric.com* and *Unpaywall*. Results indicated that articles published in open access gain higher *Altmetric Attention Scores* compared to articles behind subscription paywalls, although the difference was statistically not significant. Research outputs published in gold open access had the highest scores, followed by articles in green and then hybrid open access. Furthermore, articles by publishers with higher percentages of open access content gained higher *Altmetric Attention Scores* than articles distributed by those with medium or low percentages. In a future study additional databases could be included as well as data from years to come. Moreover, a comparable study for the field of humanities would be conceivable, including other document types such as books or contributions in anthologies as well.

Altmetrics messen, im Gegensatz zu traditionellen Metriken, den Einfluss von Forschungsergebnissen auf die breite Gesellschaft und nutzen dafür vor allem Social-Media-Plattformen als Datenquelle. In dieser Studie wurden Unterschiede in Altmetric Scores von in Open und Closed Access publizierten Artikeln deutscher Forschungseinrichtungen in den Naturwissenschaften untersucht. Hierfür wurden Daten der Jahre 2013 bis 2017 von *Web of Science*, *Altmetric.com* und *Unpaywall* gesammelt. Die Ergebnisse wiesen darauf hin, dass Artikel in Open Access höhere *Altmetric Attention Scores* erhalten als Artikel hinter Bezahlschranken. Eine statistische Signifikanz dieser Ergebnisse konnte jedoch nicht nachgewiesen werden. In Gold Open Access publizierte Forschungsergebnisse erreichten die höchsten Werte, gefolgt von in Green und Hybrid Open Access publizierten Artikeln. Zudem wiesen Artikel,

die von Verlagen mit hohen Anteilen an Open Access-Inhalten veröffentlicht wurden, höhere Scores auf als jene von Verlagen mit mittleren bis niedrigen Anteilen. In zukünftige umfassendere Studien könnten zusätzliche Datenbanken einbezogen werden sowie Daten aus den kommenden Jahren. Zudem wäre eine vergleichbare Studie für die Geisteswissenschaften denkbar, unter Einbezug weiterer Dokumententypen wie Büchern und Beiträgen in Sammelbänden.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im weiterbildenden Masterstudiengang im Fernstudium Bibliotheks- und Informationswissenschaft (Library and Information Science, M. A. (LIS)) an der Humboldt-Universität zu Berlin.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.

Acknowledgements

I would like to express my thanks to everyone who supported me during my writing of this thesis.

I thank Dr. Stephan Gauch and Prof. Vivien Petras for the supervision of this study.

I wish to thank *Altmetric.com* for providing basic data for this study free of charge for research purposes.

Very special thanks to my loved ones who supported me throughout the entire process and continuously encouraged me. My deep gratitude goes to my dear friend Cornelia Hebrank for her support with the API queries as well as her proofreading.

Contents

1. Introduction

Over the last decade, ill-applied use of traditional bibliometrics in the evaluation of research has led to critique by scientists and organizations, especially regarding the *Journal Impact Factor*. Therefore new methods were developed, the so-called *altmetrics* (short for alternative metrics or article-level metrics) which are a measure of the online attention a research output has received. The forms of attention which are assessed in these metrics are full-text downloads, views and mentions on social media platforms and in news outlets. In contrast to measures such as the *Journal Impact Factor* showing the interest in the academic community, these new metrics measure the impact research has on society at large. Furthermore, they are determining counts at the level of individual articles rather than at journal level. As the online presence and activity of scientists has increased over the last decade and is likely to continue doing so, these new metrics seem to be an effective way of assessing research impact.

A second important development over the last decade is the progressing emergence of open access publishing. In contrast to the subscription system, which is currently still widely used in some disciplines, in open access publishing access to journal articles is not paid by research institutions via subscription fees but by the authors in the form of *article processing charges* (APCs) in most cases. Articles are then freely available for download by any interested person. Due to the easier access to these articles, they might be cited more frequently than articles in toll access. This has indeed been determined by various studies and has been termed *open access citation advantage*.

However, until now the influence of open access publishing on altmetrics scores of individual articles has been investigated in only few studies which either looked at small numbers of disciplines or at specific journals, often during a short time frame. The current study aimed to broaden the view of altmetrics in natural sciences and to answer the following question:

> Do German open access journal articles in the field of natural sciences gain higher altmetrics scores than articles published behind subscription paywalls?

The study was limited to the area of natural sciences, as open access publishing is more common in these subject areas than in social sciences or humanities. In the latter subject area in many cases books are published rather than articles. Moreover, the study was limited to articles with a corresponding author residing in Germany and included universities as well as non-university research institutes and cooperations with the industry. As social media emerged in the last decade and altmetrics are a relatively new development, journal articles from 2013 to 2017 were included in this study.

In order to answer the research question, necessary data was gathered from the following three databases:

(1) *Web of Science*: From this database, general information about research outputs was downloaded including authors, title of the article, journal in which the research output has been published, addresses of the authors, *digital object identifiers* (DOIs) and the *Web of Science Research Areas* under which an article had been classified. This classification served for subdividing the natural sciences into smaller, more specific subdisciplines.

(2) *Unpaywall*: *Unpaywall* harvests data from the DOAJ (*Directory of Open Access Journals*) as well as from various journals and institutional repositories, among other sources. The open access status of each paper and of the journal in which it has been published was determined via this service. Articles were then be classified as being published in different types of open access such as gold, green, hybrid and bronze open access for this study.

(3) *Altmetric.com*: From the altmetrics aggregator *Altmetric.com* data about the online attention each article had received was gathered. This query encompassed data from various primary sources as well as the composite indicator *Altmetric Attention Score* (AAS) developed by *Altmetric.com*.

Data from the first provider was downloaded manually whereas from the other two providers it was obtained via APIs (application programming interfaces) by querying the services with DOIs. The gathered data from all three sources was put together in a SQL (structured query language) database for easier management. Data was then cleansed via SQL scripts and analyzed statistically. Furthermore, data was visualized in the form of graphical representations such as plots and histograms. Besides the main question concerning access status of articles, the influence of the following additional factors on altmetrics scores were investigated as subquestions of the main research question:

(1) Are there differences in altmetrics scores based on publication year?

(2) Are there differences in altmetrics scores based on research area?

(3) Are there differences in altmetrics scores based on publisher?

2. Literature survey

2.1. Development and use of altmetrics

Traditionally, the evaluation of research impact has been and is still based on bibliometrics, i.e. citation counts of peer-reviewed publications. The *Science Citation Index* is often used for these analyses, especially in the fields of natural and social sciences. Competing services were also created like *SciVal* based on Elsevier's *Scopus* or *Publish or Perish* and *Google Scholar Metrics*, both based on Google Scholar. The ill-applied use of these tools by institutions and organizations—concerning good practice and interpretation of metrics—has been criticized by researchers. In the *San Francisco Declaration on Research Assessment* scientists especially pointed out the deficiencies of the *Journal Impact Factor* and demanded the use of article- instead of journal-based metrics. Furthermore, they stated that alternative research outputs like data sets and software should also be considered in evaluations (DORA, 2013). In the *Leiden Manifesto* researchers proposed ten principles for responsible use of research metrics, which can also be applied to altmetrics (Bornmann & Haunschild, 2016; Hicks et al., 2015). In 2010, Jason Priem coined the term *altmetrics* (short for alternative metrics, also known as article-level metrics or social media metrics) and published the *altmetrics manifesto* concerning the use of these new metrics (Priem et al., 2010). In this manifesto they criticized the limitations of traditional metrics like peer review, citation counts and the *Journal Impact Factor* and demanded the use of new metrics including the analysis of research impact based on online activity surrounding researchers' publications. These new alternative metrics look at the broader effect of research—also on the non-scientific community—visible in the social web in the form of e.g. views, downloads and mentions referred to as usage.

Over the last years the social web has become increasingly important for researchers: for finding collaborators and staying in contact with them (Adams, 2012; Van Noorden, 2014), for disseminating their own research and discovering studies of others as well as for collaborative authoring (Rowlands et al., 2011), or for sharing open access articles via platforms like Twitter (Alperin et al., 2019). Based on Twitter data co-author networks can even be visualized via social network analysis (Robinson-García et al., 2018b).

The use of responsible metrics has also been the topic of various policy documents concerning decisions on research funding. The *National Information Standards Organization* published a recommended practice as output of its *Alternative Assessment Metrics Project* encouraging further research into altmetrics as well as proposing recommendations for higher transparency and better replicability and accuracy (NISO, 2016). In the UK, the report *The Metric Tide*—supported by the *Higher Education Funding Council for England* (HEFCE)—discussed the assessment of research output and its wider impact (Wilsdon et al., 2015). Five requirements of responsible metrics were determined: (1) robustness, (2) humility (metrics should support evaluation, not replace expert assessment), (3) transparency, (4) diversity (inclusion of factors such as variation by field) and (5) reflexivity (recognition of potential effects of indicators and regular updating of these measures). The *Research Excellence Framework* 2011 and 2014—UK's national research assessment—weighs societal impact of research at 20% in its overall evaluation criteria and defines it as follows (REF, 2011, p. 26; REF, 2014, p. 4):

> "[Impact is defined as] any effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia."

Bornmann (2013) states a similar definition of societal impact in their review of current literature.

The European Union formed an *Expert Group on Altmetrics* in 2016, which investigated the advantages and disadvantages in the use of altmetrics (Wilsdon et al., 2017). In this context the division into various levels of engagement with research is important (Haustein, 2016; European Commission, 2018): (1) accessing (a person has become aware of a research output which becomes visible as a page view or download count), (2) appraising (a research output is mentioned by a person or commented on) and (3) applying (a research output is actively used, e.g. in a lecture, presentation or publication—resulting in a citation).

Another discussion topic is whether altmetrics should be used for the evaluation of research, university rankings and for funding decisions and if they could complement traditionally used bibliometric measures. In the *Research Excellence Framework* a component of societal impact is already included, which has led to critique that such new measures need further development before being implemented into research assessment (Samuel & Derrick, 2015). Furthermore, research "varies over time and can change, positively or negatively" (Brewer, 2011, p. 256), so that a measurement of impact at one point in time does not suffice. Therefore an implementation of altmetrics into the assessment of a researchers future success might not be appropriate (Holmberg & Vainio, 2018). On the other hand, it has been concluded by many researchers that altmetrics constitute a complementary measure to bibliometrics (Bornmann, 2014; Butler et al., 2017; Haustein et al., 2015a; Thelwall et al., 2016). Therefore, altmetrics might well be included into research evaluation as additional information in the discussion of funding decisions. A merge of bibliometrics and altmetrics has been proposed as well (Taylor, 2013a; Taylor, 2013b).

Various studies have also investigated the correlation between bibliometrics and altmetrics with different results depending on the data sources. Zahedi et al. (2017) found that Mendeley readership counts could serve as a filtering tool for highly cited publications and as an early indicator of research impact. Thelwall (2018) reached similar conclusions that Mendeley readership counts correlate with later citation counts. Asemi & Heydari (2018) also determined a moderate positive correlation between readership rates on Mendeley and Research Gate and citation counts, which has also been shown by Bornmann (2015a) for Mendeley bookmark counts. Bornmann (2015b) additionally found that altmetrics from Twitter and Facebook might be applicable for the measurement of the general interest towards a publication. On the other hand, Costas et al. (2015) discovered only a weak correlation between altmetrics and citations and suggested a high validity of altmetrics only for very recent publications. Comparing altmetrics with peer review scores of societal impact, Wooldridge & King (2018) suggested that when used along with citations these new metrics could be useful as an additional measure in assessing impact. In any case, the effect of time has to be considered as with an earlier publication date more citations or a higher number of online activity around a research output can be accrued.

2.2. Data sources of altmetrics

Numerous primary and secondary data sources are used by altmetrics providers, which can be divided into various groups (Gauch & Blümel, 2016; Haustein, 2016; Sugimoto et al., 2017; Thelwall & Kousha, 2015):

Social networking Social networking sites enable users to create their own public or semi-public profile and to connect with other people on a particular service thus forming networks. These platforms are also often used by scientists to disseminate their research. Social networks for the general public are e.g. Facebook and LinkedIn whereas ResearchGate and Acacemia.edu are more focused on scholars.

Social bookmarking and reference management In social bookmarking and reference managers users can organize their various bibliographic materials, mark favorite publications and share these with others. The count of bookmarks per publication is a measure integrated by various altmetrics providers. Widely used bookmarking and reference systems are Mendeley, Zotero, CiteULike and Bibsonomy.

Social data sharing Sharing of raw data is required by various funding organizations and publishers in order to validate original research and ensure replicability. These materials can be very diverse, encompassing data sets, source code, videos and photos or presentation slides. One online repository for sharing research outputs is Figshare, which allows the dissemination of various types of material. Another, SlideShare, is especially designed for sharing presentation slides and, finally, for source code the best-known platforms are Github, Sourceforge and Bitbucket. Most of these services also allow versioning. Additionally, altmetrics providers also include data from data set repositories like Dryad and DataCite.

Video and photo sharing Videos are mostly only watched and seldom cited in publications or shared with others in a scientific context. YouTube and Vimeo are well-established video portals while Flickr is one website for sharing photos. Few scientific materials are shared on these platforms. However, the TED Talks are one format which especially focuses on science and technology and often includes academic speakers.

Mainstream media Altmetrics are also aggregated from various national and international mainstream news outlets and magazines.

Blogging Blogs are mostly written by individuals, sometimes also by small groups, and allow sharing of longer texts. Noteworthy scholarly blogs are Nature Blogs or Scientific American Blogs. There are also research blog aggregators like ResearchBlogging or ScienceSekker (Lin & Fenner, 2013). Other platforms such as WordPress and LiveJournal are not only used in the scientific field.

Microblogging Via microblogging websites scholars share recent research, often adding topical hashtags. Hashtags can be used for conducting thematic searches and staying informed in ones field of research. Undoubtedly, the most prominent platform is Twitter, allowing short posts of up to 280 characters (originally only 140 characters). Other popular platforms are e.g. Tumblr or Sina Weibo—the latter especially in China.

Wikis Wikis are platforms for collaboratively managing content in order to share knowledge and achieve a form of collective intelligence. The well-known platform Wikipedia is most widely used and includes scholarly publications as references in many articles. These are then counted as mentions by altmetrics aggregators.

Social recommending, rating and reviewing F1000Prime (Faculty of 1000) is a post-publication peer review service in the fields of biology and medicine. Peer-nominated researchers (so-called 'faculty members') review primary research they find especially interesting and tag it with terms like 'new findings', 'controversial' or 'good for teaching'. PubPeer is another platform, allowing any user to comment anonymously on scientific articles. An example for a more informal platform for rating online content and in parts scientific material is Reddit. The platform Publons allows scholars to present their peer review contributions and in doing so turns peer review into a measurable research output.

Q&A (Questions and answers) These websites cover diverse fields and are mostly self-moderating through a reputation system—users can gain reputation on the website as a reward for useful answers. An example for a popular platform is Stack Overflow, a service by Stack Exchange.

Public policy documents References to scientific research in public policy documents are also partly aggregated by altmetrics providers.

2.3. Aggregators of altmetrics

Most altmetrics aggregators are offering their data free of charge for research purposes. Altmetrics primary aggregators gather data from primary sources via different APIs (Zahedi & Costas, 2018). Additionally, some aggregators collect data not only from data sources but also from other aggregators making them secondary or tertiary aggregators (Erdt et al., 2016). In the following, the different aggregators are shortly described:

Altmetric.com *Altmetric.com*[1] is a primary aggregator. The company was founded in 2011 and gathers data from policy documents, mainstream media like news sources, reference managers, post-publication peer review services, social media, patent citations and other sources like YouTube, Wikipedia and Stack Exchange. *Altmetric.com* provides the *Altmetric Attention Score*, a quantitative measure of impact generated by a specific research output, which is mainly based on three factors: volume (how many times the research output is mentioned), sources (where the mention occurred) and authors (who mentioned

[1] https://www.altmetric.com/, visited on 03/10/2019

the output) (Altmetric.com, 2015; Elmore, 2018). *Altmetric.com* is part of the company *Digital Science*.

PLoS ALM / Lagotto PLoS ALM (Public Library of Science article-level metrics) is a primary aggregator, which launched its service in 2009, offering freely available usage statistics for every article published in PLoS. The metrics are grouped in four different categories: viewed, saved, cited and recommended. Lagotto[2] is an extended service and open source application retrieving data from PLoS and a wide range of primary sources.

Plum Analytics Plum Analytics[3] was founded in 2012 and as a secondary aggregator combines data from primary sources and PLoS ALM. Metrics are grouped into five categories: usage, captures, mentions, social media and citations. Plum Analytics is subscription-based and offers no free API, but object-level pages are freely accessible. Plum Analytics was acquired by Elsevier in 2017 and its metrics were integrated into Elsevier's products Mendeley and Scopus.

Impactstory Impactstory[4], formerly known as Total Impact, is an open source tool that enables scholars to maintain CV-like profiles showing the impact of their research. In contrast to other altmetrics providers, Impactstory calculates percentiles in order to allow comparisons between researchers across fields. Impactstory is a secondary aggregator, implementing data from *Altmetric.com* and PLoS ALM.

Webometric Analyst Webometric Analyst[5], formerly known as LexiURL, retrieves data via URL citations and title mentions from Mendeley and Bing. It also reuses data from *Altmetric.com*.

Kudos The secondary aggregator Kudos[6] aims at supporting scientists in generating higher impact and gaining visibility for their research. The service includes citation data from *Web of Science* and altmetrics from *Altmetric.com*.

Snowball Metrics The team behind Snowball Metrics[7] has the goal to create metrics that will become the global standard for institutional benchmarking. It is a tertiary aggregator and reuses data from *Altmetric.com*, Plum Analytics and Impactstory.

[2] http://www.lagotto.io/, visited on 03/10/2019
[3] https://plumanalytics.com/, visited on 03/10/2019
[4] https://profiles.impactstory.org, visited on 03/10/2019
[5] http://lexiurl.wlv.ac.uk/, visited on 03/10/2019
[6] https://www.growkudos.com/, visited on 03/10/2019
[7] https://www.snowballmetrics.com/, visited on 03/10/2019

CrossRef Event Data The service of CrossRef Event Data[8] was started in 2017 and at the time of this publication is still in beta version. It provides raw data gathered from various sources instead of calculated metrics. The raw data is available via an open API.

Altmetrics have been implemented in many publisher platforms, databases and discovery systems. One example already mentioned is the implementation of article-level metrics in PLoS. Data from *Altmetric.com* are used in the journal *Nature* and can flexibly be added to discovery systems like *Primo* and *Summon*. Plum Analytics was acquired by Elsevier in 2017, leading to the inclusion of their altmetrics in the Scopus database and in ScienceDirect.

2.4. Advantages and critique of altmetrics

While many advantages of altmetrics have been mentioned by researchers, these metrics have also been criticized in various aspects.

Altmetrics do not only measure the impact research has on the academic community, but also the broader impact beyond science, concerning the general public (Bornmann, 2014). Hence, these new metrics reveal 'hidden impacts' of research outputs which are not considered in citations such as benefits for teaching (Taylor, 2013b; Gauch & Blümel, 2016). Instead of just considering journal articles and book chapters, they also pay attention to further formats like data sets, presentation slides or source code and as such content that has not been peer-reviewed (Bornmann, 2014; Pacheco et al., 2018). Furthermore, altmetrics measure the impact of research outputs themselves and not the impact of the journal in which these have been published as metrics like the *Journal Impact Factor* do (Pacheco et al., 2018). In contrast to such citation counts, which are aggregated by subscription-based databases like *Web of Science* or *Scopus*, it is relatively easy to obtain altmetrics data from various aggregators (Bornmann, 2014). This data can be gathered very quickly after publication, unlike in the case of citations, as there is no time-delay (Bornmann, 2014). Combining several providers for large analysis may be recommendable as providers differ in their coverage of web sources, however (Ortega, 2017). Providers of altmetrics data generally aim for transparency of data sources and underlying scripts and algorithms, a goal that needs to be improved further still (Priem et al., 2010).

One of the major concerns around the use of altmetrics is data quality and integrity. The variety of different data collection methodologies regarding APIs, doc-

[8] https://www.eventdata.crossref.org/, visited on 03/10/2019

ument types and unique identifiers has been criticized and the importance of regular update frequencies has been stressed (Zahedi & Costas, 2018). This is due to the use of different APIs depending on data aggregators as well as the tracking of different document types, journals and publishers and dependence on unique identifiers like DOIs (digital object identifiers), PMIDs (PubMed IDs) or arXiv IDs. A big concern are the already heterogenous data sources as the web is very dynamic and replicability of results cannot be guaranteed (Haustein, 2016; Zahedi & Costas, 2018). The lack of transparency concerning the underlying methodology for data collection and processing by altmetrics providers is a point of critique (Zahedi & Costas, 2018). Various data aggregators are already addressing this problem, though. In order to allow comparisons across fields and across time, normalization and standardization of definitions and terminology have to be applied, as clear definitions of terminology—specifying terms such as view—as well as measurement standards—concerning different handling mentions based on extent—are rarely used (Bornmann, 2014). Manipulation is another concern around altmetrics as higher altmetrics scores can easily be achieved by researchers using fake accounts or bots (Bornmann, 2014). However, in some cases countermeasures to prevent and discover systematic abuse have already been implemented (Gordon et al., 2015). It has also been shown that the dependence on unique identifiers like DOIs leads to a certain bias as use of unique identifiers is not common in all parts of the world (Alperin, 2015). The dependency on a small number of data providers—especially *Altmetric.com* as one of the few primary aggregators—leads to an almost monopolistic position similar to the case of the *Science Citation Index* (Haustein, 2016).

Concerning manipulation, it has to be kept in mind that while altmetrics are more susceptible to this issue, it also occurs with citations. Common techniques for inflating citation counts are e.g. salami-publishing, self-plagiarism, honorary authorship, authorship for sale, self-citation and citation cartels (Haustein & Larivière, 2015; Haustein, 2016). Furthermore, a point of critique in common for both bibliometrics and altmetrics is the lack of assessing the quality of research outputs instead of only their metrics (Bornmann & Haunschild, 2018a; Bornmann & Haunschild, 2018b).

Another important fact is that citation and usage differ significantly, with the former being public and the latter being rather private (Moed & Halevi, 2016). Furthermore, reading and citing populations are not congruent and can be divided into user groups like researchers, practitioners, under- and postgraduates and the interested public (Kurtz & Bollen, 2010; Nicholas et al., 2005). Additionally, there are different levels of engagement with research output—as already mentioned above—and downloading an article does not mean that it will also be fully read or actually used (Moed & Halevi, 2016). Moreover, other factors like humorous or declarative titles also affect the sharing of research via social media (Di Girolamo & Reynders, 2017; Holmberg, 2015). Similarly, novel or emotional topics or ones especially suitable for a wider audience generally receive higher online attention (Holmberg & Vainio, 2018).

A noteworthy type of measurement is the *Altmetric Attention Score* (AAS), a composite indicator of online attention and output that has been critically assessed by various studies. This score is a weighted count including multiple online data sources, which are evaluated through an automated algorithm (Elmore, 2018). Bringing various sources together in such a single score can be problematic as dependencies exist between the underlying components that may influence or cancel out each other (Gumpenberger et al., 2016). Moreover, online reference management systems like Mendeley are not included as a data source in the AAS (Gumpenberger et al., 2016). Mukherjee et al. (2018) also discourage the use of the AAS for group comparisons due to measurement inconsistencies. Nevertheless, this score is used by many researchers to monitor the online attention around their research output (Holmberg & Vainio, 2018), and after evaluating and adjusting the weighting schemes could be integrated into traditional metrics (Huang et al., 2018). In their study, Black et al. (2018) concluded that the *Altmetric Attention Score* might correlate with future citation counts.

Overall, altmetrics are a useful tool to measure the impact research receives in social media and society at large and could be used as an additional measure to complement the traditionally used bibliometrics (Haustein et al., 2015b; Thelwall & Kousha, 2015).

2.5. Emergence and definitions of open access

In 2002 open access was defined as making content *free to read* as well as *free to reuse* by the *Budapest Open Access Initiative* (BOAI)[9]. This is equivalent to the Creative Commons License 'CC-BY' (Creative Commons, 2013). In 2003 the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* and the *Bethesda Statement on Open Access Publishing* built on the following definition of open access formulated in the BOAI:

> "[Open access to literature means] free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them

[9] `https://www.budapestopenaccessinitiative.org/read`, visited on 03/10/2019

> as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

Bearing this in mind, open access (OA) can be subdivided into four major groups (Archambault et al., 2014; Piwowar et al., 2018):

Gold OA articles that are published in a pure open access journal that is listed in the *Directory of Open Access Journals* (DOAJ).

Green OA articles published in a subscription journal but self-archived in an open repository, e.g. in the form of a preprint in the repository arXiv or as the original publication in an institutional or domain-specific repository. Institutional repositories are listed in the *Directory of Open Access Repositories* (OpenDOAR) or the *Registry of Open Access Repositories* (ROAR).

Hybrid OA articles published in a subscription journal that are free to read in exchange for the payment of article processing charges (APCs) paid by the authors.

Bronze OA articles that are free to read on the publisher's websites or in repositories but have no clear licenses displayed.

There are further definitions dividing open access into *Libre OA* (research outputs that are free to read and reuse) and *Gratis OA* (articles that are only free to read). Additionally, publishers offer *Delayed OA*, making articles freely available after an embargo period. In the case of public health emergencies like the 2013-2016 Ebola virus outbreak or the 2015 Zikus virus outbreak, publishers make articles freely available for a certain time in *Transient OA* (Littler et al., 2017). In the case of *Black* or *Rogue OA*, articles are shared illegally on pirate sites like Sci-Hub. What has been criticized in the case of *Hybrid OA* is the so-called *double dipping* meaning that universities or libraries have to pay two times for access to articles, first in the form of APCs and then again as subscription fees (Pinfield et al., 2016). In contrast, a not-for-profit model of open access is *Diamond OA*, which encompasses free publication of research outputs without subscription fees or APCs but prohibits commercial use of these materials (Fuchs & Sandoval, 2013). This model can be funded either through public funding models or through mandatory funding by universities, faculties, research councils or higher education and research assessment institutions (Fuchs & Sandoval, 2013).

In general, the aim is to reorganize the current business model and payment streams and transform the current system into a more sustainable one based on open access (Pinfield, 2015; Schimmer et al., 2015). Many funding organizations like the *Deutsche Forschungsgemeinschaft* (DFG) have also implemented policies into their guidelines that research output of projects funded by them should be made available as open access. In addition, the DFG is conducting the program *Open Access Publizieren* promoting open access in German research (DFG, 2017; DFG, 2018). A similar transnational endeavour is *PlanS*, an initiative of the consortium *cOAlition S*[10], which includes numerous major European funding organizations and research agencies. Archambault et al. (2014) showed that open access had gained ground since 1996, and that in 2014 over 50% of the articles published from 2007 to 2012 could be downloaded freely as open access versions.

2.6. Open access citation advantage

Several studies have shown that articles published in open access gain higher citation counts than articles published behind subscription paywalls. Archambault et al. (2014) conducted an investigation at European and world level, examining articles published between 2008 and 2013 in different types of open access. They found a citation advantage of 40% for open access publications and a disadvantage of not publishing in open access of 27%, which was the measured average of all fields. The biggest advantage was gained by publishing in green open access whereas gold open access was more difficult to assess due to the trend for gold open access journals to be smaller and younger. Swan (2010) compiled studies conducted between 2001 and 2010 on the open access citation advantage, the results of which indicated that open access publications achieve two to three times higher citation rates than articles in subscription journals (Tunger, 2017). Another, smaller study based on articles published between 2007 and 2009 reinforced the existence of an open access citation advantage (Archambault et al., 2016). For preprints, Davis & Fromerth (2006) showed that articles gained 35% more citations by being made available via arXiv. Two other studies found a citation advantage of 17% and 18% respectively for either self-archiving in green open access or hybrid open access, with this effect seemingly declining over the years 2004 to 2007 for green open access (Davis, 2009; Piwowar et al., 2018). In their study, Eysenbach (2006) concluded that open access publications are twice as likely to be cited as articles in toll access and that non-open

[10] `https://www.coalition-s.org/`, visited on 03/10/2019

access articles are more likely to remain uncited. Overall, citation advantages have been determined in various studies for publishing in green, gold and hybrid open access.

Three postulates explaining the advantage of publishing in open access have been proposed (Davis & Fromerth, 2006; Kurtz et al., 2005; Norris et al., 2008; Swan, 2010; Wren, 2005):

Open access postulate Articles published in open access are available to a wider audience, read by more researchers and in turn cited more often.

Early access postulate Open access articles have a citation advantage as they are often published earlier than those in subscription journals, therefore gaining additional time in press and a higher chance of being cited.

Self-selection bias postulate / Quality advantage Authors are preferentially making their best, most citable articles available in open access or well-known authors are using open access as a means to publish their research.

There are also studies refuting the existence of an open access citation advantage. Björk & Solomon (2012) showed that open access journals reach a similar scientific impact as subscription journals in the field of biomedicine when considering discipline, journal age and country of the publisher. However, gold open access articles were cited more often compared to other open access variants. Pacheco et al. (2018) found that type of access is no relevant criterion for a successful paper as the most important journals in different fields will be available to scientists through their university or library and restricted access to the less relevant journals is negligible.

In relation to altmetrics, several studies have determined an open access citation advantage. Alhoori et al. (2015) found higher altmetrics scores for open access articles even though the significance decreased when also considering journal, publication year and citation count, while Davis et al. (2008) did not determine an open access citation advantage in the field of physiology. They noted an increase of 89% for full-text downloads and 23% more unique visitors. Investigating two journals in the field of physics, Wang et al. (2015) also found a higher number of downloads for open access articles over a longer period of time compared to non-open access publications, indicating that open access articles might receive higher attention over longer time spans. Moreover, gold open access publications receive more mentions than green open access ones and are covered to a higher extent by *Altmetric.com* (Robinson-García et al., 2018a). Finally, for the fields of intensive care medicine and anesthesia, Black et al. (2018) showed higher *Altmetric Attention Scores* for open access articles compared to ones published in paywalled journals.

2.7. Context of the current study

The occurrence and magnitude of an open access advantage has mostly been assessed for citation counts. There are few studies measuring this effect for the relatively new altmetrics, and these studies either have focused on a small number of journals in a specific discipline or investigated only high-ranked journals. Additionally, most studies looked at articles published only during a short time frame. Davis et al. (2008) collected data from 11 journals in the field of physiology in the first half of 2007. Alhoori et al. (2015) assessed articles from various fields published between 2010 and 2014 in 23 journals that were highly ranked by Google Scholar. Black et al. (2018) collected data from Medline (PubMed) for the fields of intensive care medicine and anesthesia published in 2015. Wang et al. (2015) assessed articles published in the journal *Nature Communications* in 2012 and 2013, and Robinson-García et al. (2018a) investigated two review journals in the field of physics, *Physical Review B* and *Physical Review X*, including data from 2011 to 2018. All these studies investigated between 1,800 and 40,000 journal articles respectively.

To add onto the scope of these earlier studies, the current study aimed to investigate data encompassing articles from all fields of natural sciences published between 2013 and 2017, amounting to around 65,000 records in total with around 18,000 of these records containing *Altmetric.com* scores. The investigation included articles independent of any specific journals in which they were published but focused on the open access status of the individual articles themselves as well as their journals' open access status. Classification of articles into different subfields of natural sciences were done in accordance with the *Web of Science Research Areas* assigned to each article in the *Web of Science Core Collection*. This study tried to assess a possible open access advantage of articles published in gold, green, hybrid and bronze open access compared to articles published in subscription journals by looking at the altmetrics scores gained by either group. The aim of this study was to gain a broader view of altmetrics in natural sciences.

3. Methods and implementation of the study

3.1. Coverage of databases queried

The data for the study was gathered from three databases, which were queried for different types of data: relevant articles were identified via *Web of Science Core Collection*, then their access status was checked via *Unpaywall* and their altmetrics scores collected from *Altmetric.com*.

Web of Science Core Collection was chosen for bibliographic data as this database offers high quality metadata of articles published in peer-reviewed journals. The *Web of Science Research Areas* were used for the classification of articles into different research areas as these are subdivided into various disciplines for the life sciences, biomedicine, physical sciences and technology. Archambault et al. (2009) showed the number of articles retrieved via *Web of Science* and via *Scopus* to be similar in their study. In this study data was only collected from *Web of Science*. As *Web of Science* applies selection criteria for including journals in the database, the analyzed sample in this study is not exhaustive. Collection of further data from additional databases like *Scopus* with subsequent cleansing of duplicate entries would have resulted in a more complete sample.

The availability of an open access version of articles was checked through *Unpaywall*, which gathers data from the DOAJ (*Directory of Open Access Journals)* as well as from publisher databases and various repositories. The open access status offered by the *Web of Science* was not used as it has been found to be less detailed by Robinson-García et al. (2018a). Only data listed in the `best_oa_location` field was analyzed. As the underlying algorithm prioritizes publisher-hosted content over content on repositories, green open access was probably underestimated in this study. If multiple sources are found for an article, *Unpaywall* lists further results in the `oa_locations` field, which was not analyzed in this study. Furthermore, delayed open access was not discernable as data snapshots were taken in this study.

Altmetrics data was gathered from the primary aggregator *Altmetric.com* as this service offers counts for different social media platforms, reader counts for online reference management systems and a composite indicator, the *Altmetric Attention Score* (AAS).

Data was analyzed through various statistical tests, chosen based on data characteristics, and results were displayed in plots. The study focused on access status of articles, but additionally included factors like publication year, research area and publisher.

3.2. Overview of procedure

The following steps were taken to gather and cleanse the data (see Figure 1). Individual steps were described in more detail in the following sections.

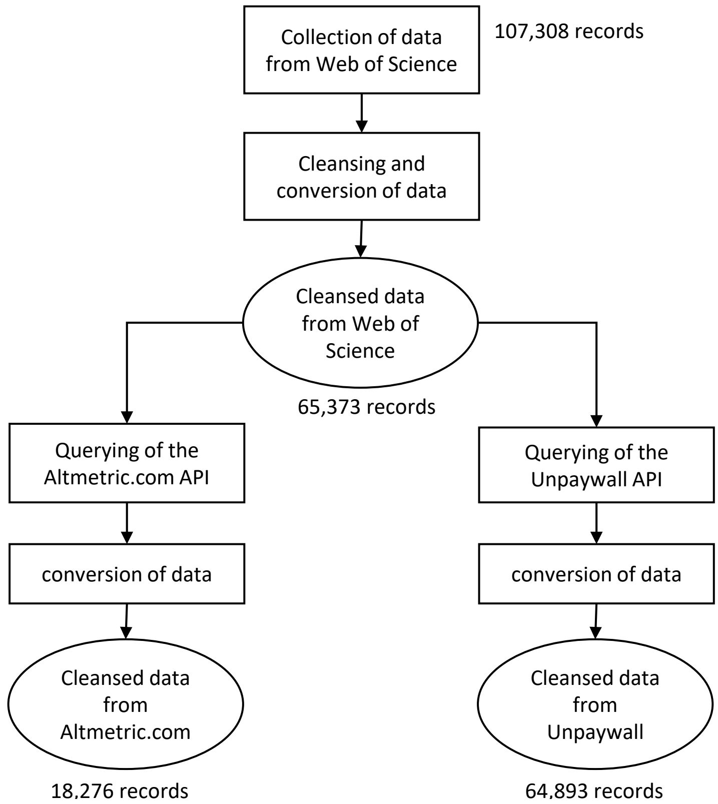

Figure 1: General procedure and steps in gathering and cleansing data

3.3. Collection and cleansing of data from *Web of Science*

Data was collected from the *Web of Science Core Collection* via the following query (with DO = DOI; CU = Country/Region; SU = Research Area; DT = Document Type; PY = Year Published), including data indexed up to 23/02/2019:

```
(DO=(10.*)) AND
(CU=(Germany)) AND
(SU=(Life Sciences Biomedicine OR Physical Sciences OR
Acoustics OR Computer Science OR Energy & Fuels OR
Engineering OR Imaging Science & Photographic Technology OR
Instruments & Instrumentation OR Materials Science OR
Mechanics OR Metallurgy & Metallurgical Engineering OR
Microscopy OR Nuclear Science & Technology OR
```

```
Remote Sensing OR Robotics OR Spectroscopy))
AND (DT=(Article))
AND (PY = (2013-2017))
```

This encompassed articles

(1) published in a journal,

(2) based in the areas of life sciences, biomedicine, physical sciences and selected technological disciplines,

(3) published between 2013 and 2017,

(4) with a German address as reprint address (data includes university research, non-university research as well as articles published in cooperation with the German industry),

(5) with a DOI as identifier.

The following indices were included into the search through the *Web of Science Core Collection*:

- *Science Citation Index Expanded* (SCI-EXPANDED)—1945-present
- *Social Sciences Citation Index* (SSCI)—1956-present
- *Arts & Humanities Citation Index* (A&HCI)—1975-present
- *Book Citation Index – Science* (BKCI-S)—2013-present
- *Book Citation Index – Social Sciences & Humanities* (BKCI-SSH)—2013-present
- *Emerging Sources Citation Index* (ESCI)—2015-present

This query resulted in 107,308 articles fulfilling the criteria. Results were downloaded through the interface provided by the *Web of Science* database. For this, articles had to be added to a so-called *Marked List* and downloaded in packages of up to 500 results, for which additionally desired fields were selected on top of the normally provided ones—namely addresses, document type, publisher info and research areas. The data was downloaded in the format 'Windows, tab delimited, utf-8'. Collected data was then read into a local SQL database provided by a MariaDB service[11] (version 10.3.13) and cleansed through a number of SQL queries:

[11] https://mariadb.org, visited on 03/10/2019

(1) Articles without a reprint address or corresponding author were excluded (368 records):

```
delete
from wos_dataset
where repr_address = ""
```

(2) Articles which did not have the publication type *Journal* were excluded (3,271 records: 882 type *Book*, 2,389 type *Series* = book chapter):

```
delete
from wos_dataset
where publication_type != "J"
```

(3) Articles which did not have the document type *Article* were excluded (7,081 records: 7 type *Data Paper*, 2 type *Retracted Publication*, 7,072 type *Proceedings Paper*):

```
delete
from wos_dataset
where doc_type != "Article"
```

(4) Articles which did not include 'Germany' in their reprint address were excluded (31,214 records):

```
delete
from wos_dataset
where repr_address not like "%Germany%"
```

Furthermore, one article was indexed two times in *Web of Science*, under the same DOI and title but with different entries in the journal field. This mistake was corrected through checking the actual data by entering the DOI in a browser and subsequently deleting the false record. After data cleansing, 65,373 records remained and were used for gathering additional data from *Altmetric.com* and *Unpaywall*.

3.4. Collection and cleansing of data from *Altmetric.com*

Altmetric.com offers a *Details Page API* for retrieving count data from their database. This service has no rate limit when used with an API key, which was kindly provided by *Altmetric.com* for this study. The API was queried on 23/02/2019 with the DOIs of the 65,373 *Web of Science* records using the following query structure with the respective **DOI_name** inserted and a key **xxx** given:

`https://api.altmetric.com/v1/doi/`**DOI_name?key=xxx**

The API response included the requested data in JSON format (*JavaScript Object Notation*). A Java-based tool was used to automatically extract and convert JSON raw data and finally save converted data to the SQL database (the complete source code can be accessed via the link given in the footnote)[12]. For each dataset, the JSON string was split at each comma and all containing curly brackets were excluded. The resulting smaller substrings were saved in an array (in the form `field-name : field-value`), which was then converted into an attribute map (composed of attribute-value pairs) through the following steps: Using regular expressions, substrings were split at the colon and double quotes were removed, resulting in the first part being the field name and the second part the corresponding value. As the DOI can contain a colon in some cases, this was checked before splitting the substring and in these cases a workaround was used to avoid splitting DOIs. Finally, a database object was created from the attribute map with all fields of interest mapped to a suitable data type—such as string, integer, date. These database objects were then saved to a separate altmetrics table in the SQL database. Desired fields included into the database as columns were the following (fields as defined by *Altmetric.com*[13]):

- `cited_by_x_count`—mention count breakdown, with `x` being number of:
  - `delicious`—Delicious users
  - `fbwalls`—Facebook accounts
  - `feeds`—blogs
  - `forum`—internet forums users
  - `gplus`—Google+ users
  - `linkedin`—LinkedIn users
  - `msm`—news outlets
  - `peer_review_sites`—peer review sites
  - `pinners`—Pinterest users
  - `policies`—policy sources
  - `qs`—questions, answers or comments on Stack Exchange sites

[12] Hebrank, Cornelia (2019): *Metrics-Converter-Server*, https://gitlab.com/CHebrank/metrics-converter-server, visited on 03/10/2019 (logic for converting data to be found in `DataConverter.java`)

[13] http://api.altmetric.com/docs/call_citations.html, visited on 03/10/2019

  - rdts—Reddit users
  - rh—research highlight platforms
  - tweeters—Twitter users
  - videos—YouTube channels
  - weibo—Sina Weibo users
  - wikipedia—pages on Wikipedia
- cited_by_posts_count—total number of posts
- readers_x—number of readers on platform x:
  - citeulike
  - connotea
  - mendeley
- readers_count—total reader count
- cohorts_x—number of people mentioning this article who are members of x, with x being:
  - com—science communicators
  - doc—practitioners
  - pub—the public
  - sci—research scientists
- score—Altmetric Attention Score
- history_x—what the cited article's Altmetric score would be if you only looked at posts from the past x, with x being:
  - 1y—one year
  - 6m—six months
  - 3m—three months
- added_on—date article was first tracked by Altmetric
- last_updated—time the data for this resource was last updated

While gathering data, errors were caused by non-utf-8 symbols as the *Altmetric.com Details Page API* also allows utf-8mb4 symbols in the tq-field (descriptive phrases for an article). The 55 records affected by this were therefore added manually to the SQL database after deleting these symbols. In total, altmetrics data for 18,276 articles was collected, which corresponds to 28.0% of the records collected from the *Web of Science*. For the remaining 47,089 DOI queries (72.0%), no altmetrics data was provided by *Altmetric.com*.

3.5. Collection and cleansing of data from *Unpaywall*

Unpaywall is an open service that provides information about access status for up to 100,000 records per day via an API. Queries containing the DOIs of the 65,373 *Web of Science* records were sent to the API on 23/02/2019 using the following syntax with the respective **DOI_name** inserted and an email address **xxx** given:

`https://api.unpaywall.org/v2/`**DOI_name**`?email=`**xxx**

The API response included the requested data in JSON format. Similar to the process for *Altmetric.com* data (see subsection 3.4) the conversion tool was used to automatically extract, convert and save datasets. However, in the *Unpaywall* datasets contents between square brackets were additionally removed as they were not of interest. Again, resulting attribute maps were mapped to database objects for a separate unpaywall table. Desired fields included into the database were the following (as defined by *Unpaywall*[14]):

- `best_oa_location`—The 'best' location is determined using an algorithm that prioritizes publisher-hosted content first (hybrid or gold), then versions closer to the version of record next (published version over accepted version), and more authoritative repositories last.

- `data_standard`—Indicates the data collection approaches used for this resource. Possible values are '1' (first-generation hybrid detection using only data from the Crossref API to determine hybrid status) or '2' (second-generation hybrid detection using additional sources and checking all publishers for hybrid; this results in about 10 times as much being classified as hybrid).

- `host_type`—type of host that serves this OA location, either 'publisher' or 'repository'

[14] https://unpaywall.org/data-format, visited on 03/10/2019

- `license`—The license under which this copy is published, can either be Creative Commons licenses, publisher-specific licenses or ‘implied-oa’ when there is evidence that an OA license of some kind was used, but it is not reported directly on the webpage at this location.
- `version`—The content version accessible at this location, can either be ‘submitted version’ (not yet peer-reviewed), ‘accepted version’ (peer-reviewed, but lacks publisher-specific formatting) or ‘published version’ (is the version of record).
- `is_oa`—‘true’ if there is an OA copy of this resource.
- `journal_is_in_doaj`—‘true’ if this resource is published in a journal indexed by DOAJ.
- `journal_is_oa`—‘true’ if this resource is published in a full OA journal.
- `published_date`—Date this resource was published as reported by the publisher.
- `updated`—Time the data for this resource was last updated.

Only data listed in the `best_oa_location` field was analyzed, further data listed under `oa_locations` was not included into the analysis. In total, the access status of 64,893 articles could be determined, which corresponds to 99.3% of the records collected from *Web of Science*. For the remaining 480 articles (0.7%), a ‘not found’ response was returned for the *Unpaywall* queries. Regarding the *Altmetric.com* dataset, for 18,256 records (99.9%) the *Unpaywall* API supplied results. For the focused analysis of altmetrics data as well as the analysis of differences in altmetric scores, this final dataset of 18,256 records was used. For the broader analysis of data, the full datasets of 65,373 records (from *Web of Science*) and 18,276 records (from *Altmetric.com*) were used respectively in order to gain an overview of data.

3.6. Definitions of investigated factors

Inspired by the *Web of Science Research Areas*—based on the categorization of the journal—, the total of 107 research areas was divided into the following broader disciplines:

- **Life Sciences**: Agriculture; Anatomy & Morphology; Anthropology; Behavioral Sciences; Biodiversity & Conservation; Biophysics; Biotechnology & Applied Microbiology; Cell Biology; Developmental Biology; Entomology; Evolutionary Biology; Fisheries; Food Science & Technology; Forestry; Genetics &

Heredity; Marine & Freshwater Biology; Mathematical & Computational Biology; Microbiology; Mycology; Paleontology; Parasitology; Physiology; Plant Sciences; Zoology

- **Biomedicine**: Allergy; Anesthesiology; Cardiovascular System & Cardiology; Critical Care Medicine; Dentistry, Oral Surgery & Medicine; Dermatology; Emergency Medicine; Endocrinology & Metabolism; Gastroenterology & Hepatology; General & Internal Medicine; Geriatrics & Gerontology; Health Care Sciences & Services; Hematology; Immunology; Infectious Diseases; Integrative & Complementary Medicine; Legal Medicine; Medical Ethics; Medical Informatics; Medical Laboratory Technology; Neurosciences & Neurology; Nursing; Nutrition & Dietetics; Obstetrics & Gynecology; Oncology; Ophthalmology; Orthopedics; Otorhinolaryngology; Pathology; Pediatrics; Pharmacology & Pharmacy; Psychiatry; Public, Environmental & Occupational Health; Radiology, Nuclear Medicine & Medical Imaging; Rehabilitation; Reproductive Biology; Research & Experimental Medicine; Respiratory System; Rheumatology; Sport Sciences; Substance Abuse; Surgery; Toxicology; Transplantation; Tropical Medicine; Urology & Nephrology; Veterinary Sciences; Virology

- **Physical Sciences**: Astronomy & Astrophysics; Crystallography; Electrochemistry; Geochemistry & Geophysics; Geology; Mathematics; Meteorology & Atmospheric Sciences; Mineralogy; Mining & Mineral Processing; Oceanography; Optics; Physical Geography; Polymer Science; Thermodynamics; Water Resources

- **Technology**: Acoustics; Imaging Science & Photographic Technology; Instruments & Instrumentation;Mechanics; Metallurgy & Metallurgical Engineering; Microscopy; Nuclear Science & Technology; Remote Sensing; Robotics; Science & Technology - Other Topics; Spectroscopy

One article can be attributed to various defined research areas in *Web of Science*. Duplications inside these four defined research areas were resolved so that each article only counted once per area.

Additionally, single research areas more than 1,000 records or a percentage of over 5% respectively of the total records in the *Altmetric.com* dataset were handled as individual research areas. These were the following:

- **Biochemistry & Molecular Biology**

- **Environmental Sciences & Ecology**

- **Life Sciences & Biomedicine - Other Topics**
- **Chemistry**
- **Physics**
- **Computer Science**
- **Energy & Fuels**
- **Engineering**
- **Materials Science**

Access status based on data from *Unpaywall* was defined as follows:

- **gold open access**:
  `is_oa` = True
  `journal_is_oa` = True
  `host_type` = publisher
  `license` = ‘cc’
- **green open access**:
  `is_oa` = True
  `host_type` = repository
  `license` = ‘cc’
- **hybrid open access**:
  `is_oa` = True
  `journal_is_oa` = False
  `host_type` = publisher
  `license` = ‘cc’
- **bronze open access**:
  `is_oa` = True
  `license` != ‘cc’
- **closed access**:
  `is_oa` = False
- **unclear access status**:
  articles not listed in the *Unpaywall* database

The publication year of the printed version was provided in the *Web of Science Core Collection* sample.

Different branch names or abbreviations of the same publisher were combined into one common name. For an analysis of the *Altmetric Attention Score* based on the percentage of open access content by publisher, publishers were divided into three groups: low percentage of open access content (0-33.3%), medium percentage (33.4-66.6%) and high percentage (66.7-100%). Publishers that were represented with only 1 or 2 records in the *Altmetric.com* dataset were excluded from this analysis (49 publishers with a total of 64 articles; 18,192 articles by 95 publishers remained for analysis).

3.7. Statistical analysis

Statistical analysis was performed in the programming language R (version 3.5.2) (R Core Team, 2017) using the development environment RStudio (version 1.1.463) (RStudio Team, 2016). Use of statistical tests was determined based on the following rules:

- Whether data was normally distributed was determined using the Anderson-Darling normality test (AD) from the R package `nortest` (Gross & Ligges, 2015). If the p-value was above 0.05 ($p > 0.05$), normality of data was assumed. With a p-value of below 0.05 ($p < 0.05$), data was assumed to be not normally distributed.

- Whether variances of data were homogeneous or heterogeneous was determined using the Fligner-Killeen-test (FK). For $p < 0.05$ heterogeneity of variances was assumed, whereas for $p > 0.05$ homogeneity of variances was assumed.

- Comparison of two groups with not normally distributed data and homogeneous or heterogeneous variances was conducted using the non-parametric Mann-Whitney-Wilcoxon-U-test (MWU). For $p < 0.05$ a significant difference between the two groups was assumed, whereas for $p > 0.05$ no significant difference was assumed.

- Comparisons of more than two groups were conducted using the Kruskal-Wallis-test (KW). If this test resulted in $p < 0.05$, significant differences between groups were assumed, whereas with $p > 0.05$ no significant differences were assumed. If differences were found, a subsequent pairwise comparison between all groups was conducted using the Mann-Whitney-Wilcoxon-U-test

(MWU). This test was applied with the Bonferroni correction in order to correct for multiple testing errors as with multiple comparisons the likelihood of falsely rejecting the null hypothesis increases (Type I error). Similar to the Kruskal-Wallis-test, significant differences were assumed based on whether the p-value was above or below 0.05.

All values were denoted as mean ± standard error (SE), which was calculated from the variance (s^2) and the sample size (n) with x_i being one observation in the dataset and $\overline{x}$ being the arithmetic mean of all observations, as follows:

$$SE = \sqrt{\frac{s^2}{n}}$$

$$s^2 = \frac{\sum_{n=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Plots were done using the packages `ggplot2` (Wickham, 2016) and `scales` (Wickham, 2018) with the `viridis` color palette (Garnier, 2018). Whiskers of boxplots are defined as $1.5 \cdot$ IQR (interquartile range).

3.8. Caveats

This study did not exclusively include university research, but also non-university research and articles published in cooperation with the German industry, as clear classification of every organization to one group is rather complex. Gathered data from *Web of Science* was selected only when one of the corresponding authors had an affiliation with seat in Germany. *Web of Science* merges similar branch names of institutions in the *Organization Enhanced* field. This possibility to retrieve data was not used as it would have been necessary to compile a complete list of all higher education institutions and non-university research institutions, which would have been prone to errors and incompleteness. Instead, another approach was taken using the program OpenRefine[15] in order to get a rough estimate of how many articles in gathered samples had institutions of the German industry as sole corresponding authors. Reprint address fields containing multiple reprint addresses were split up into several columns subdividing the field at periods, which are used by *Web of Science* to divide addresses in this field. The single columns were then searched for the word ‘Germany’ and all addresses not containing this word were discarded. Afterwards, the columns were again merged to one single column now containing only addresses based in Germany. This field was then filtered searching for higher

[15] http://openrefine.org/, visited on 03/10/2019

education institutions and non-university research institutions. These records were then discarded leaving only research published by solely the German industry. Filter terms included the most commonly used designations for German higher education institutions as well as important non-university research institutions (see Table 1):

Table 1: Filter terms for estimating the percentage of the German industry as sole corresponding author in German research

Filter terms	
Leibnizinst	Leibniz Inst
Max Planck	Max Plank
MPI *case sensitive*	Fraunhofer
Helmholtz	KIT *case sensitive*
Karlsruhe Inst	Karlsruher Inst
DESY	Synchrotron
Forschungszentrum Julich	FZ Julich
Res Ctr Julich	Fernuniv
Univ *case sensitive*	Uni *case sensitive*
Forschungsinstitut	Forschungsinst
Hsch *case sensitive*	Fachhsch *case sensitive*
Fachhochschule	FH *case sensitive*
TU *case sensitive*	TH *case sensitive*
TUM *case sensitive*	LMU *case sensitive*
FAU *case sensitive*	Fak *case sensitive*
Lehrstuhl	Chair *case sensitive*

case sensitive indicates that searches were conducted in this mode. When using short names such as 'MPI' or 'Univ' spaces and or commas before and after letters were included respectively into the search in order to avoid searching for parts of other words.

This method resulted in a rough overstated estimate only as it was also prone to some errors:

- Many universities abbreviate their names—examples are TUM, LMU, FAU. Not all of these abbreviations are well-known.
- Designations like 'institute' or 'department' cannot be assigned clearly to one group since universities as well as the industry include this word in their names.
- Sometimes it is not clear if an institution belongs to a higher research institution as in the case of hospitals which can be part of universities but can also be private clinics.
- Correction for spelling mistakes are not possible for all names, although some common ones have been included like 'Max Plank' (correctly Max Planck).

Via this method it was estimated that 6.4% (4213 records) of the data retrieved from *Web of Science* (65,373 records in total) and 4.2% (772 records) of the records retrieved from *Altmetric.com* (18,276 records in total) had a member of the German industry as sole corresponding author. It was decided to leave these records in the data sample as no completely clean assignation of all records to one group—university, non-university, industry—could be achieved.

Another aspect that has to be kept in mind is that altmetrics data was gathered from only one aggregator, *Altmetric.com*, but other data aggregators do exist such as *Plum Analytics* or *Lagotto*. This study was conducted looking only at data from one aggregator as the inclusion of various sources for data retrieval and subsequent data cleansing and merging would have exceeded the scope of and time for this study.

A multivariate analysis of the *Altmetric Attention Score* based on publication year, access status of articles and percentage of open access content by publisher was not conducted. For normally distributed data in this case a three-way factorial ANOVA (Analysis of variance) would have been used. As data was not normally distributed and had heterogeneous variances, a non-parametric approach was necessary instead. The `ARTool` package allows for such analysis -with its command `anova.art` which conducts an *Aligned Rank Transform Analysis of Variance* (Wobbrock et al., 2011; Kay & Wobbrock, 2019). However, conditions for this test were not met as F-values were not all close to 0. To the author's knowledge, no other equally appropriate test for non-parametric data with three independent variables—each with multiple factors—and one dependent variable is available at the time of this study.

3.9. Additional information about data availability

Raw data from *Web of Science*[16], *Altmetric.com*[17] and *Unpaywall*[18] is not allowed to be freely redistributed. Aggregated data of figures is included in Appendix A.

[16] `https://clarivate.com/wp-content/uploads/dlm_uploads/2018/12/End-User-Terms.pdf`, visited on 03/10/2019

[17] `https://drive.google.com/file/d/17xNAui1ZAN3LgCzx59Te1DYmAfGfaK-Q/view`, visited on 03/10/2019

[18] `https://unpaywall.org/legal/terms-of-service`, visited on 03/10/2019

4. Results and interpretation

4.1. Broad analysis of data

In the following, the distribution of data from *Web of Science* and *Altmetric.com* is shown for various factors such as

- publication year (as defined in subsection 3.6),
- access status (as defined in subsection 3.6, based on data from *Unpaywall*),
- research area (as defined in subsection 3.6) and
- publisher (as defined in subsection 3.6).

4.1.1. Publication year

The number of articles published was found to have increased over the years 2013 to 2017, which was mirrored in the investigated datasets (see Figure 2). Additionally, a higher percentage of articles from 2016 and 2017 were indexed by *Altmetric.com* compared to research outputs from previous years.

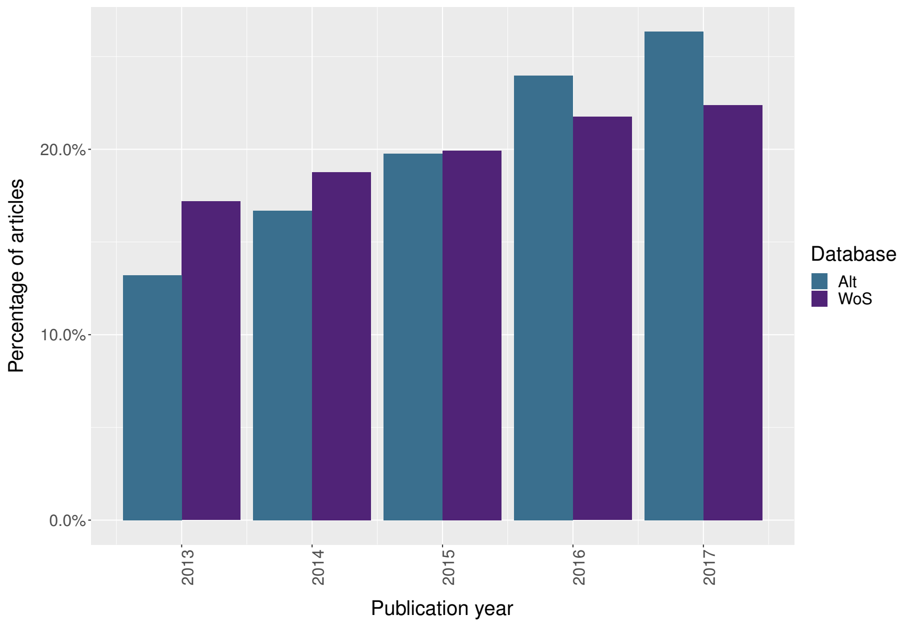

Figure 2: Percentage of articles per publication year (2013 to 2017), based on the datasets retrieved from *Web of Science* (WoS) and *Altmetric.com* (Alt) (WoS: n = 65,373; Alt: n = 18,276)

4.1.2. Access status

Over this range of years the percentage of articles published in different open access variants—gold, green, hybrid and bronze—increased from 20.4% to 30.7% (as indexed by *Web of Science*; see Figure 3). Most articles published in open access were available as bronze open access versions (up to 17.5% per year in the *Web of Science* dataset and up to 34.2% per year in the *Altmetric.com* dataset). Compared to closed access versions, more open access articles were indexed by *Altmetric.com*, which became evident by comparing their data to the original data retrieved from *Web of Science*. Nonetheless, the percentage of articles in closed access is rather high with 68.5% in 2017 to 78.9% in 2013, although it shows a diminishing trend.

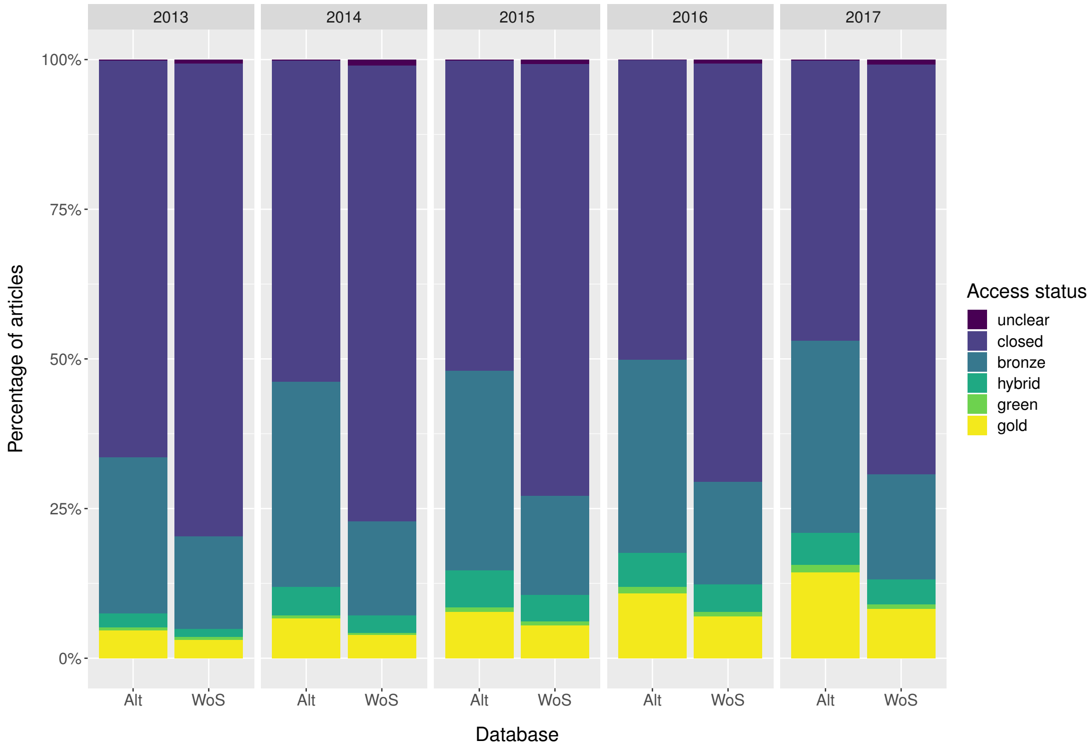

Figure 3: Percentage of articles per access status, based on the datasets retrieved from *Web of Science* (WoS) and *Altmetric.com* (Alt)
(WoS: n = 65,373; Alt: n = 18,276)
Access status (as defined in subsection 3.6): *gold* = gold open access, *green* = green open access, *hybrid* = hybrid open access, *bronze* = bronze open access, *closed* = closed access, *unclear* = unclear access status (not listed in *Unpaywall* database)

4.1.3. Research area and access status

Most articles in the *Web of Science* (WoS) and *Altmetric.com* (Alt) datasets were published in the areas of Materials Science (WoS: 38.7%; Alt: 41.0%), Engineering

(WoS: 35.0%; Alt: 24.3%), Physics (WoS: 22.7%; Alt: 30.9%) and Technology (WoS: 33.1%; Alt: 30.9%) (note that assignment of articles to more than one discipline was possible; see subsection 3.6 for a definition of listed research areas) (see Figure 4). Indexing of articles in *Altmetric.com* was not evenly distributed, leading to an underrepresentation of some disciplines, i.e. Engineering and overrepresentation of others such as Physics.

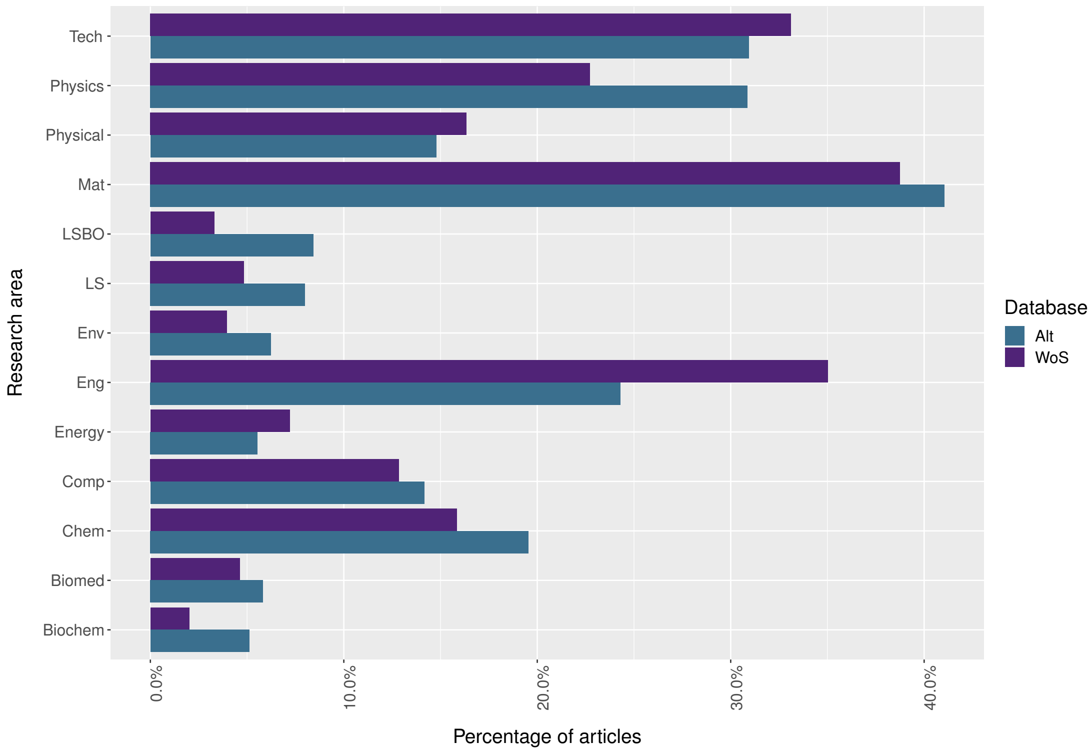

Figure 4: Percentage of articles per research area, based on the datasets retrieved from *Web of Science* (WoS) and *Altmetric.com* (Alt)

(WoS: n = 65,373; Alt: n = 18,276)

Note: Assignment of articles to more than one discipline is possible.

Listed research areas are (as defined in subsection 3.6): 1) Biochem (Biochemistry & Molecular Biology), 2) Biomed (Biomedicine), 3) Chem (Chemistry), 4) Comp (Computer Science), 5) Energy (Energy & Fuels), 6) Eng (Engineering), 7) Env (Environmental Sciences & Ecology), 8) LS (Life Sciences), 9) LSBO (Life Sciences & Biomedicine - Other Topics), 10) Mat (Materials Science), 11) Physical (Physical Sciences), 12) Physics, 13) Tech (Technology)

A higher percentage of altmetrics data was available for open access articles compared to closed access ones (WoS: closed access in average 68.3%, open access in average 31.1%; Alt: closed access in average 53.6%, open access in average 46.4%) (see Figure 5). Proportions between certain open access status—gold, green, hybrid, bronze—were similar for both datasets in all disciplines.

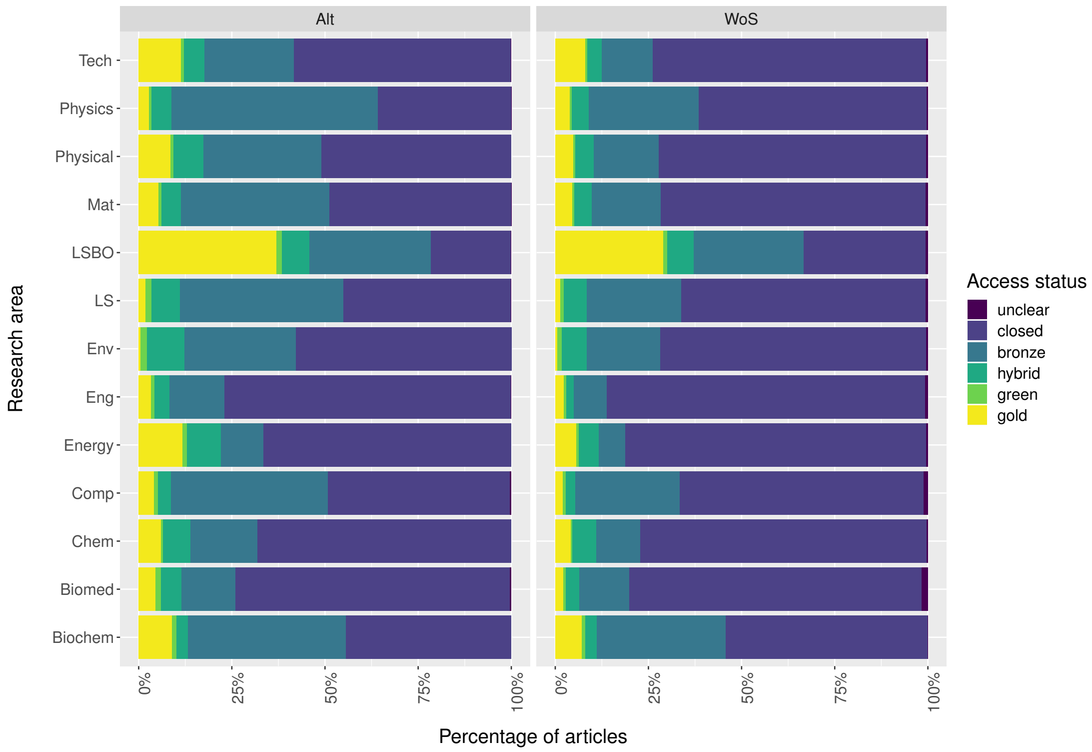

Figure 5: Percentage of articles per research area and access status, based on datasets retrieved from *Web of Science* (WoS) and *Altmetric.com* (Alt) (WoS: n = 65,373; Alt: n = 18,276)
Note: Assignment of articles to more than one discipline is possible.
Access status (as defined in subsection 3.6): *gold* = gold open access, *green* = green open access, *hybrid* = hybrid open access, *bronze* = bronze open access, *closed* = closed access, *unclear* = unclear access status (not listed in *Unpaywall* database)

4.1.4. Publisher and access status

For the *Altmetric.com* dataset, 144 different publishers were identified with eleven major publishers distributing 82.4% of all articles published in the years investigated. For the *Web of Science* dataset, 327 publishers were recorded, with eleven of the major publishers distributing 80.3% of articles, ten of which were also major in the *Altmetric.com* dataset (see Figure 6). Of those publishers *Elsevier* had by far the highest percentage of articles distributed (WoS: 29.1%; Alt: 18.5%). Additionally, *Wiley*, *Springer*, the *American Physical Society* and the *American Chemical Society* published substantial percentages of articles, ranging between 5% and 15%.

MDPI (*Multidisciplinary Digital Publishing Institute*) exclusively published articles in gold open access (see Figure 7). *Oxford University Press* and the *American Physical Society* distributed a majority of articles in bronze open access (76.8% and 70.3% respectively), whereas the *Royal Society of Chemistry* published a high percentage of articles in hybrid open access (31.4%). The percentage of closed ac-

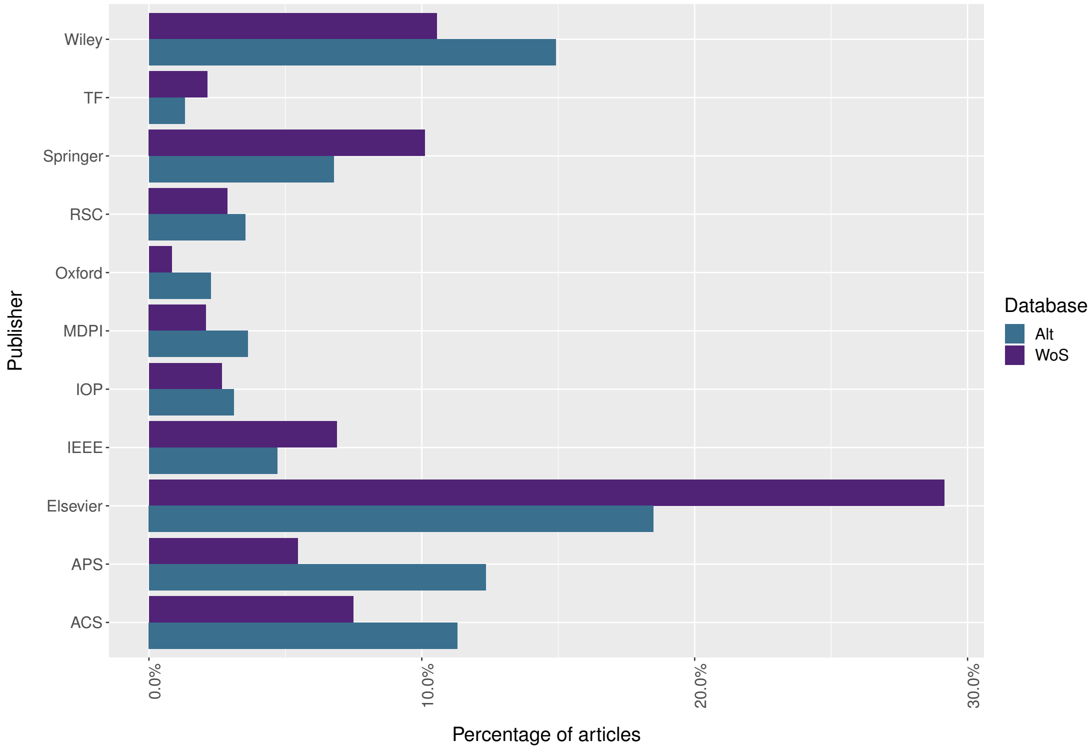

Figure 6: Percentage of articles distributed by major publishers, based on the dataset retrieved from *Altmetric.com* with their respective percentages as found on *Web of Science*

(WoS: n = 65,373; Alt: n = 18,276)

Listed publishers are (as defined in subsection 3.6): 1) Wiley, 2) TF (Taylor & Francis), 3) Springer, 4) RSC (Royal Society of Chemistry), 5) Oxford (Oxford University Press), 6) MDPI (Multidisciplinary Digital Publishing Institute), 7) IOP (Institute of Physics), 8) IEEE (Institute of Electrical and Electronics Engineers), 9) Elsevier, 10) APS (American Physical Society), 11) ACS (American Chemical Society) cess varied strongly between publishers—ranging from 20.7% of *Oxford University Press*'s publications to between 84.4% and 89.0% for major publishers like *Elsevier*, *Wiley* as well as the *American Chemical Society* and *Taylor & Francis*. These proportions based on the *Web of Science* dataset were mirrored in the *Altmetric.com* sample.

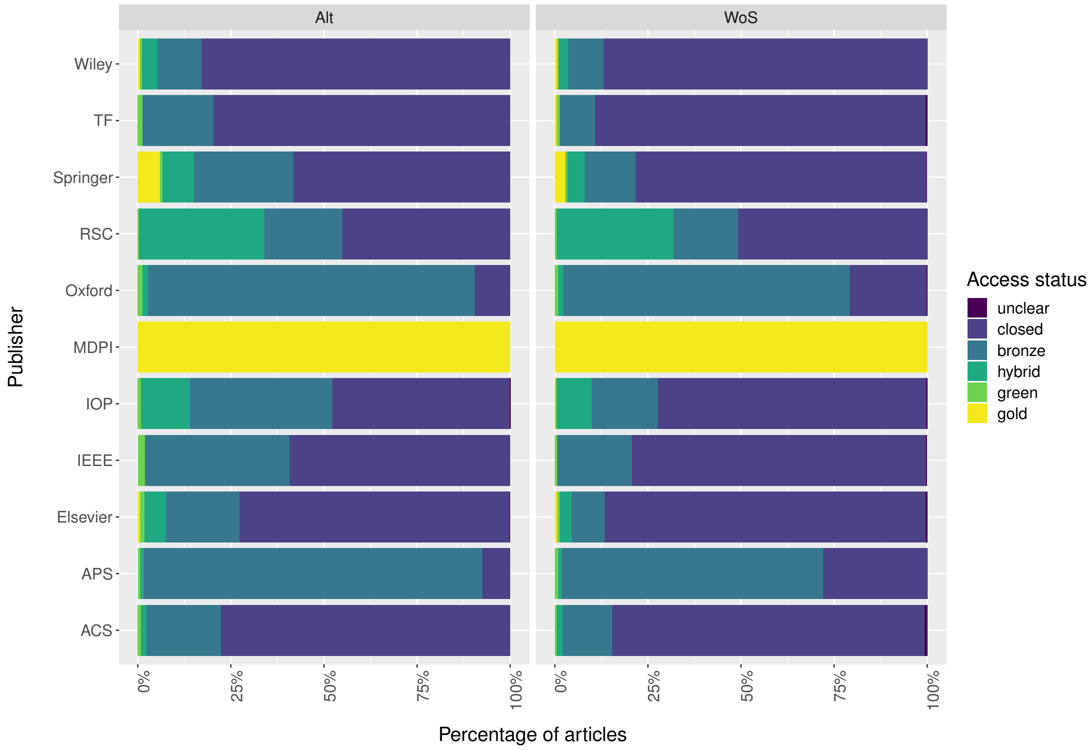

Figure 7: Percentage of articles distributed by major publishers in different access status, based on the datasets retrieved from *Web of Science* and *Altmetric.com*

(WoS: n = 65,373; Alt: n = 18,276)

Listed publishers are (as defined in subsection 3.6): 1) Wiley, 2) TF (Taylor & Francis), 3) Springer, 4) RSC (Royal Society of Chemistry), 5) Oxford (Oxford University Press), 6) MDPI (Multidisciplinary Digital Publishing Institute), 7) IOP (Institute of Physics), 8) IEEE (Institute of Electrical and Electronics Engineers), 9) Elsevier, 10) APS (American Physical Society), 11) ACS (American Chemical Society) Access status (as defined in subsection 3.6): *gold* = gold open access, *green* = green open access, *hybrid* = hybrid open access, *bronze* = bronze open access, *closed* = closed access, *unclear* = unclear access status (not listed in *Unpaywall* database)

4.2. Focused analysis of altmetrics data

Subsequently, a closer analysis of the *Altmetric.com* dataset was conducted, including data about online mentions as well as about the people mentioning research outputs, and readers counts on bookmarking and reference management systems. However, the main focus of this study lies on the *Altmetric Attention Score*, which will be further investigated in the next subsection.

Concerning online mentions, the count of total posts was mostly composed of tweeters (67.0%) as well as news outlets (8.0%), and to a smaller extent of Facebook accounts (4.5%) and blogs (2.0%) (see Figure 8). The counts are weighted and then integrated into the composite indicator *Altmetric Attention Score*. Mentions of research outputs in news outlets and in blogs count a multiple of tweeters with the latter in turn counting a multiple of Facebook comments (see Table 2)[19]. On average, research outputs were mentioned in 4.90 (± 0.26) posts with 3.27 (± 0.17) tweeters sharing information about articles.

Table 2: Weighted counts of the *Altmetric Attention Score*

Data source	Weighted count
News	8
Blogs	5
Wikipedia	3
Policy Documents (per source)	3
Patents	3
Twitter	1
Sina Weibo	1
F1000/Publons/Pubpeer	1
Google+	1
Open Syllabus	1
LinkedIn	0.5
Facebook	0.25
YouTube	0.25
Q&A	0.25
Reddit/Pinterest	0.25

Research was mostly mentioned by the public (66.3%) as well as science communicators (26.8%), and only to a smaller degree by research scientists (5.2%), with practitioners contributing 1.7% to the total amount (see Figure 9). On average, an article was shared by two members of the public and up to one member of the

[19] https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-attention-score-calculated, visited on 03/10/2019

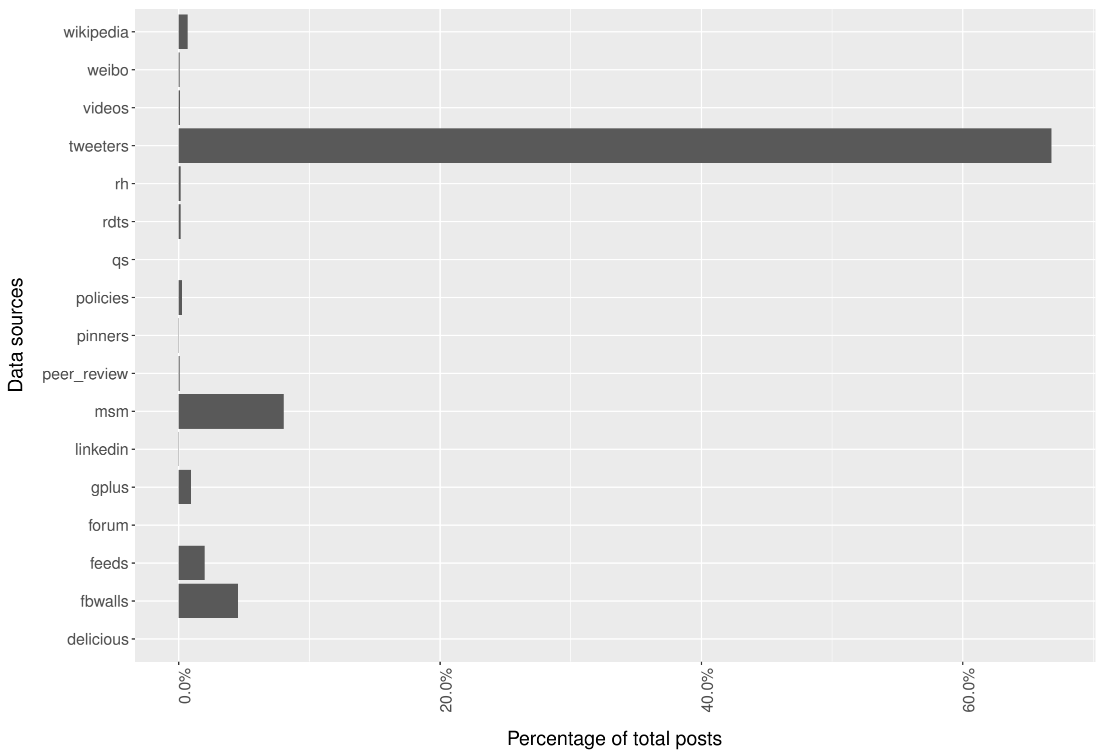

Figure 8: Percentage of total posts by data source, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Listed data sources are (as defined in subsection 3.6): 1) wikipedia (pages on Wikipedia), 2) weibo (Sina Weibo), 3) videos (YouTube channels), 4) tweeters (Twitter users), 5) rh (research highlight platforms), 6) rdts (Reddit users), 7) qa (questions, answers or comments on Stack Exchange sites), 8) policies (policy sources), 9) pinners (Pinterest users), 10) peer_review (peer review sites), 11) msm (news outlets), 12) linkedin (LinkedIn users), 13) gplus (Google+ users), 14) forum (internet forums users), 15) feeds (blogs), 16) fbwalls (Facebook accounts), 17) delicious (Delicious users) other cohorts (the public: 2.17 ± 0.12; science communicators: 0.88 ± 0.05; research scientists: 0.17 ± 0.01; practitioners: 0.05 ± 0.01).

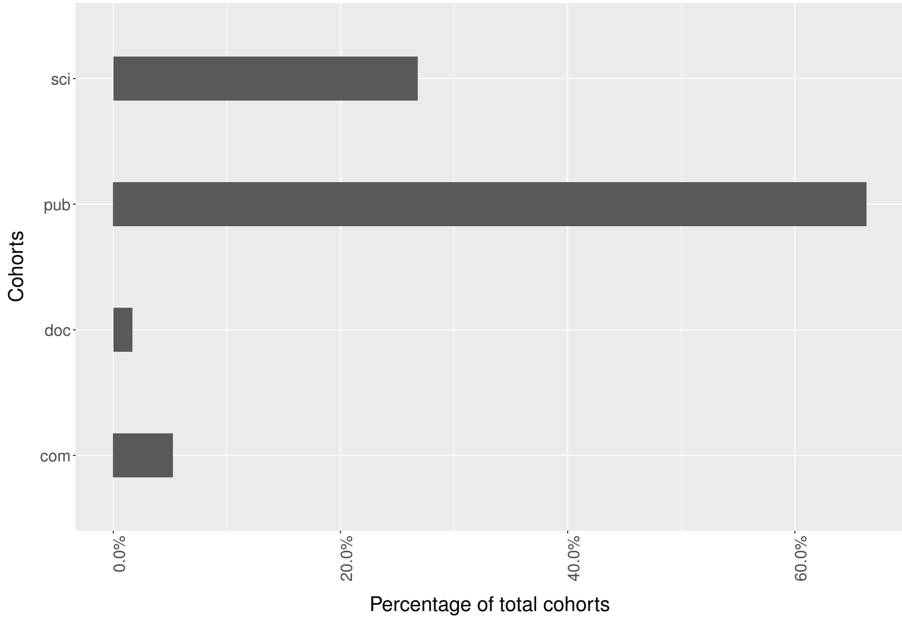

Figure 9: Percentage of total cohorts by cohort, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Listed cohorts are (as defined in subsection 3.6): 1) sci (science communicators), 2) pub (the public), 3) doc (practitioners), 4) com (research scientists)

Readers counts were almost exclusively due to listings as Mendeley bookmarks, with listings in CiteULike and Connotea only contributing 0.003% and 0.0003% respectively to the total reader count. On average, a single article was saved 35.12 (± 0.57) times in Mendeley.

The average *Altmetric Attention Score* was 5.94 (± 0.21) with values ranging from 0 up to 1974.29. The two articles with the highest values (the only ones over 1000) were either highly mentioned in news outlets (> 200 times) or tweeted often (> 2500 tweeters). Over the first year, the *Altmetric Attention Score* of articles seemed to rise slowly from an average of 0.06 (± 0.01) after three months up to 0.41 (± 0.03) after one year.

4.3. Analysis of *Altmetric Attention Score*

In the following, *Altmetric Attention Scores* are represented based on the investigated factors publication year, access status, research area and publisher (all as defined in subsection 3.6).

4.3.1. Publication year

Data of the *Altmetric Attention Score* was not normally distributed (AD $p < 2.2e\text{-}16$) and variances in relation to publication year were heterogeneous (FK, $p < 2.2e\text{-}16$). *Altmetric Attention Scores* were highest in the year 2017 with a mean of 6.99 ($\pm$ 0.61) and lowest in the year 2015 with a mean of 4.72 ($\pm$ 0.24) (see Figure 10). This difference could be due to the fact that there were multiple high outliers.

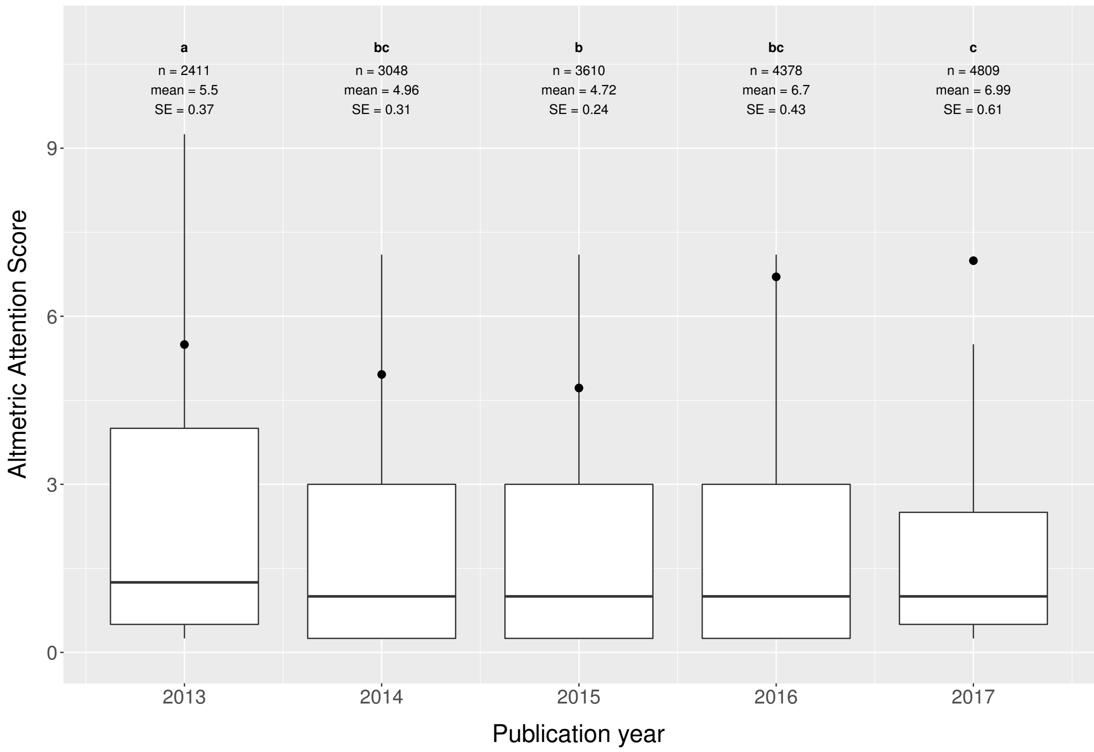

Figure 10: *Altmetric Attention Score* per publication year, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Note: Only values between 0 and 10 are shown, outliers are not displayed but were included into the analysis.

Sample sizes (n), means and standard errors (SE) are denoted for each publication year. Significant differences between publication years are indicated through differing letters. Means are displayed as black dots.

4.3.2. Access status

Variances of the *Altmetric Attention Score* in relation to access status (categories: open, closed) were homogeneous (FK, $p = 0.052$). *Altmetric Attention Scores* of open access articles did not differ significantly from those of closed access ones, although the former had a higher mean of 7.08 (± 0.37) compared to the latter with a mean of 4.90 (± 0.22) (MWU $p = 0.7392$) (see Figure 11). An analysis excluding the two highest outliers (*Altmetric Attention Scores* higher than 1000) revealed similar results (means: 6.72 ± 0.26 and 4.90 ± 0.22; MWU $p = 0.7392$).

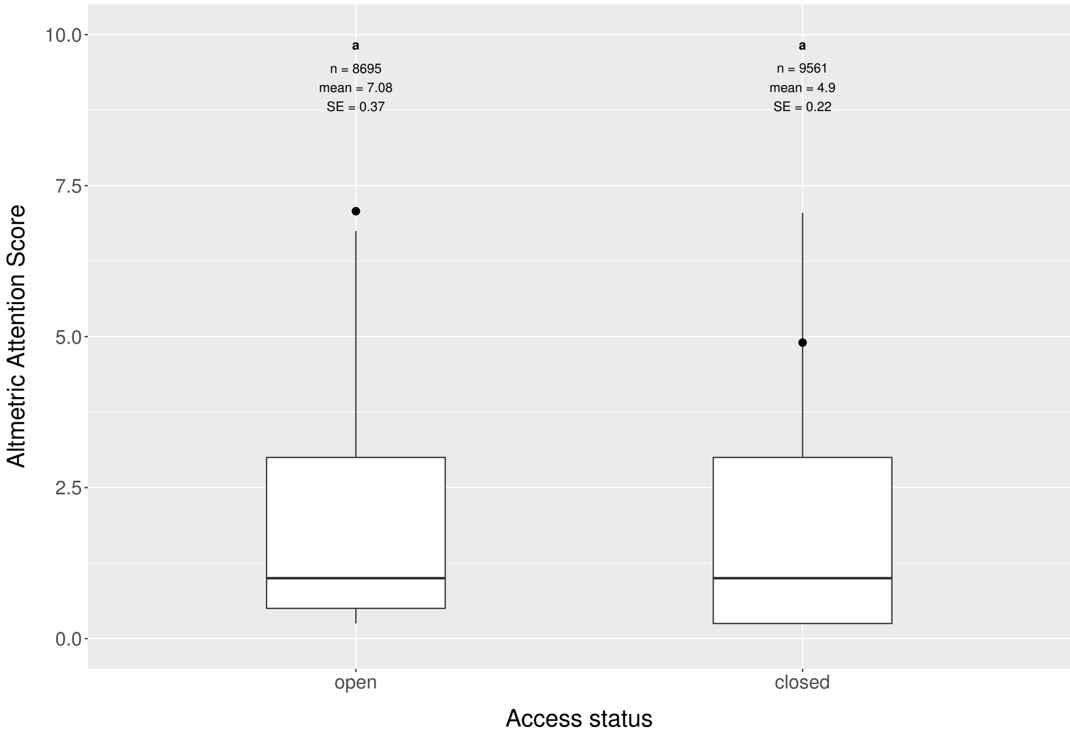

Figure 11: *Altmetric Attention Score* for open and closed access, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Note: Only values between 0 and 7.5 are shown, outliers are not displayed but were included into the analysis.

Sample sizes (n), means and standard errors (SE) are denoted for each access status. Significant differences between access status are indicated through differing letters. Means are displayed as black dots.

Variances of the *Altmetric Attention Score* in relation to access status (categories: gold, green, hybrid, bronze, closed) were heterogeneous (FK, $p < 2.2e\text{-}16$). Articles published in gold open access had significantly higher *Altmetric Attention Scores* than those published in other access status except for green open access (KW $p <$ 2.2e-16; MWU $p < 3.1e\text{-}06$) (see Figure 12). Green open access articles had similar *Altmetric Attention Scores* compared to gold and hybrid open access (MWU p =

1.0), although the sample size for this group was the smallest and the standard error was relatively high (n = 166; SE = 3.01). *Altmetric Attention Scores* of research outputs distributed in bronze and closed access differed significantly from articles of all the other groups (MWU $p < 0.019$). Excluding the two very high values for the *Altmetric Attention Score* among the gold open access articles, comparisons between groups revealed the same results. Only the mean of *Altmetric Attention Scores* for gold open access decreased from 13.31 (± 1.53) to 11.57 (± 0.84).

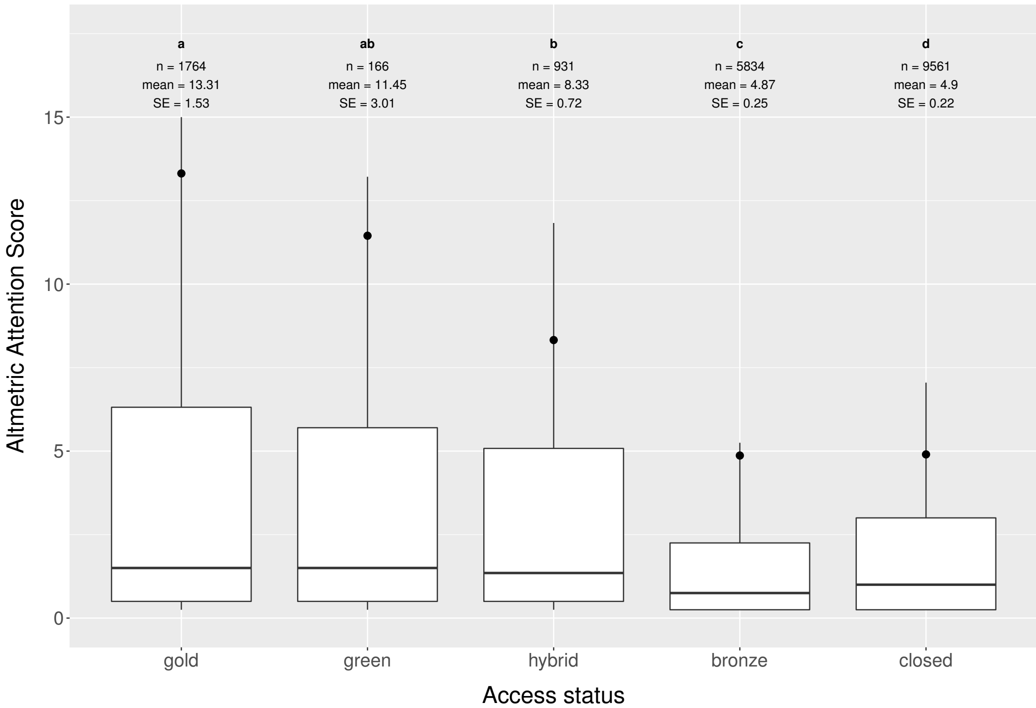

Figure 12: *Altmetric Attention Score* per access status, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Note: Only values between 0 and 15 are shown, outliers are not displayed but were included into the analysis.

Sample sizes (n), means and standard errors (SE) are denoted for each access status. Significant differences between access status are indicated through differing letters. Means are displayed as black dots.

Access status (as defined in subsection 3.6): *gold* = gold open access, *green* = green open access, *hybrid* = hybrid open access, *bronze* = bronze open access, *closed* = closed access

4.3.3. Research area and access status

Variances of the *Altmetric Attention Score* in relation to research area were heterogeneous (FK, $p < 2.2e\text{-}16$). The research areas *Life Sciences & Biomedicine - Other Topics* and *Environmental Sciences* had the significantly highest values with means of 22.88 (± 1.86) and 14.85 (± 1.33) respectively (KW p < 2.2e-16; MWU p < 6.3e-05) (see Figure 13). Articles published in the disciplines *Biomedicine* and *Computer Science* had the lowest values with means of 3.92 (± 0.55) and 3.94 (± 0.29) respectively.

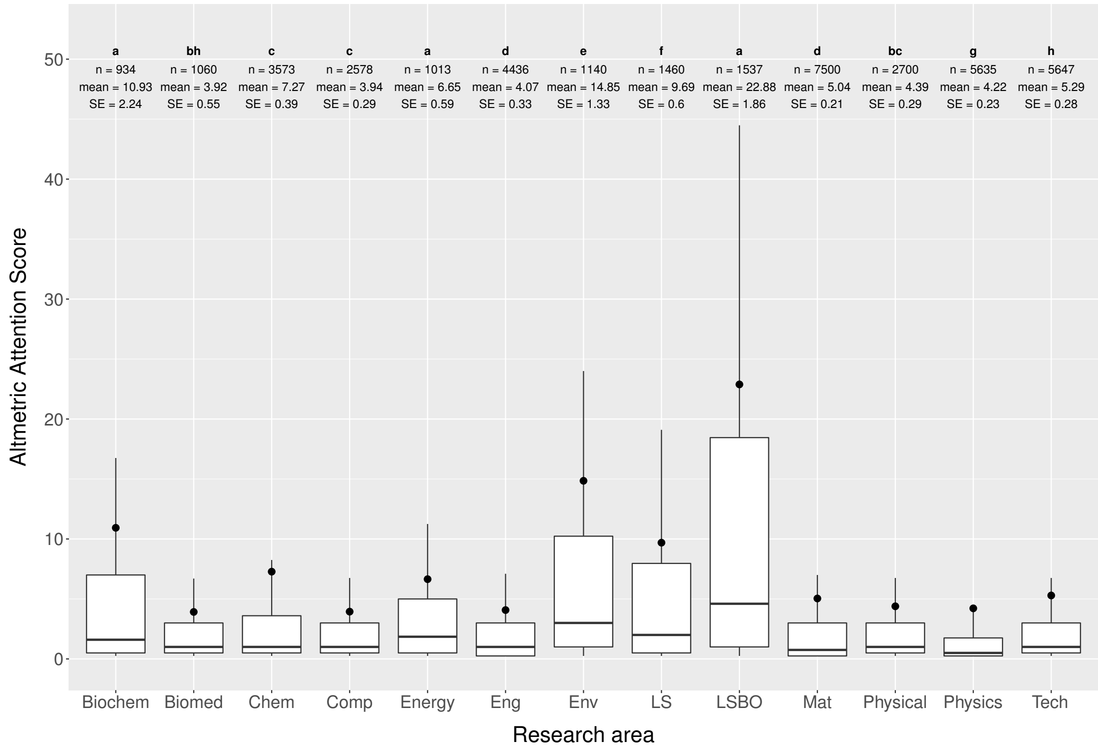

Figure 13: *Altmetric Attention Score* per research area, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Note: Only values between 0 and 45 are shown, outliers are not displayed but were included into the analysis.

Sample sizes (n), means and standard errors (SE) are denoted for each research area. Significant differences between research areas are indicated through differing letters. Means are displayed as black dots.

Listed research areas are: 1) Biochem (Biochemistry & Molecular Biology), 2) Biomed (Biomedicine), 3) Chem (Chemistry), 4) Comp (Computer Science), 5) Energy (Energy & Fuels), 6) Eng (Engineering), 7) Env (Environmental Sciences & Ecology), 8) LS (Life Sciences), 9) LSBO (Life Sciences & Biomedicine - Other Topics), 10) Mat (Materials Science), 11) Physical (Physical Sciences), 12) Physics, 13) Tech (Technology)

4.3.4. Publisher and access status

When only investigating eleven major publishers, variances of the *Altmetric Attention Score* in relation to publisher were heterogeneous (FK p < 2.2e-16). Articles published by *Oxford University Press* had the significantly highest *Altmetric Attention Scores* with a mean of 8.54 (± 0.858) (KW $p < 2.2e\text{-}16$; MWU $p < 2e\text{-}16$) whereas research outputs published by the *American Physical Society* had the significantly lowest *Altmetric Attention Scores* with a mean of 1.43 (± 0.13) (MWU $p < 2e\text{-}16$) (see Figure 14), although for both publishers the majority of their published contents was available in bronze open access.

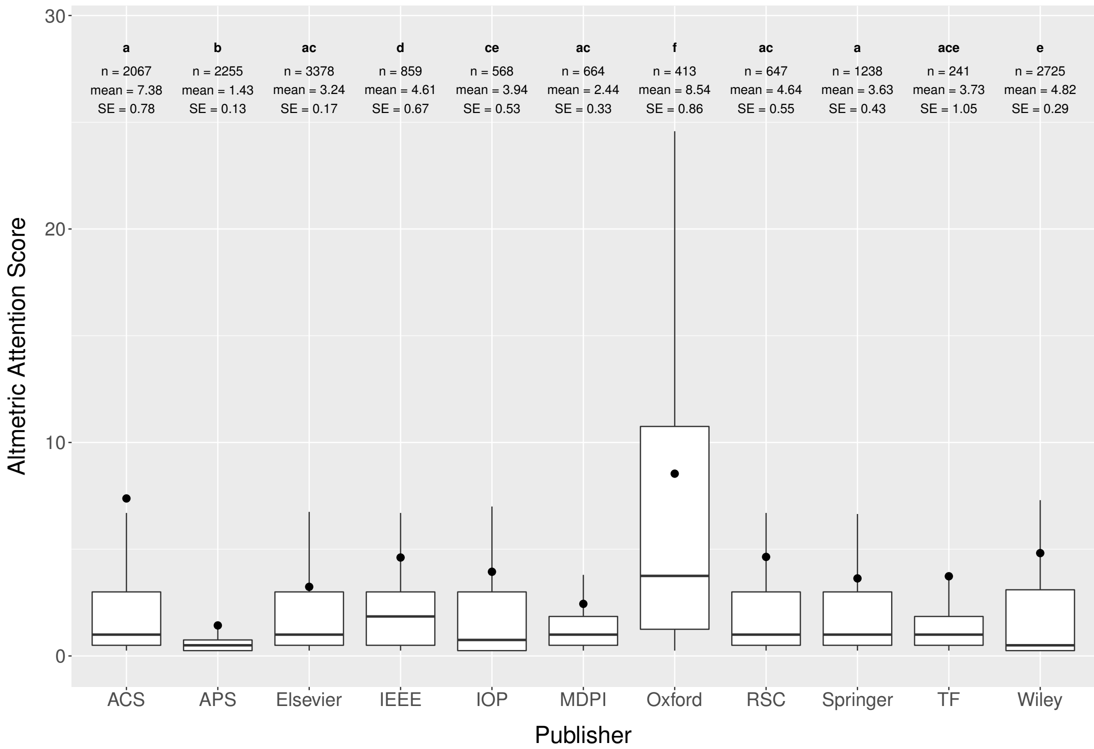

Figure 14: *Altmetric Attention Score* based on major publishers, based on the dataset retrieved from *Altmetric.com* (n = 18,256)

Sample sizes (n), means and standard errors (SE) are denoted for each publisher. Significant differences between publishers are indicated through differing letters. Means are displayed as black dots.

Note: Only values between 0 and 25 are shown, outliers are not displayed but were included into the analysis.

Listed publishers are: 1) ACS (American Chemical Society), 2) APS (American Physical Society), 3) Elsevier, 4) IEEE (Institute of Electrical and Electronics Engineers), 5) IOP (Institute of Physics), 6) MDPI (Multidisciplinary Digital Publishing Institute), 7) Oxford (Oxford University Press), 8) RSC (Royal Society of Chemistry), 9) Springer, 10) TF (Taylor & Francis), 11) Wiley

When dividing all 95 publishers into categories displaying the percentage of open access content—low ((0-33.3%), medium (33.4-66.6%), high (66.7-100%)—variances of *Altmetric Attention Score* were heterogeneous (FK $p = 7.6e\text{-}4$). *Altmetric Attention Scores* in the category medium differed significantly from the other two categories (KW p = 4.632e-12; MWU p < 2.1e-07), whereas the categories low and high did not differ significantly from each other (MWU p = 0.83) (see Figure 15). Nonetheless, articles distributed by publishers with a high percentage of their content in open access received higher *Altmetric Attention Scores* with a mean of 8.39 (± 0.60) than those in the category medium with a mean of 5.62 (± 0.33) and the latter in turn receiving higher scores than those in the category low with a mean of 4.78 (± 0.23).

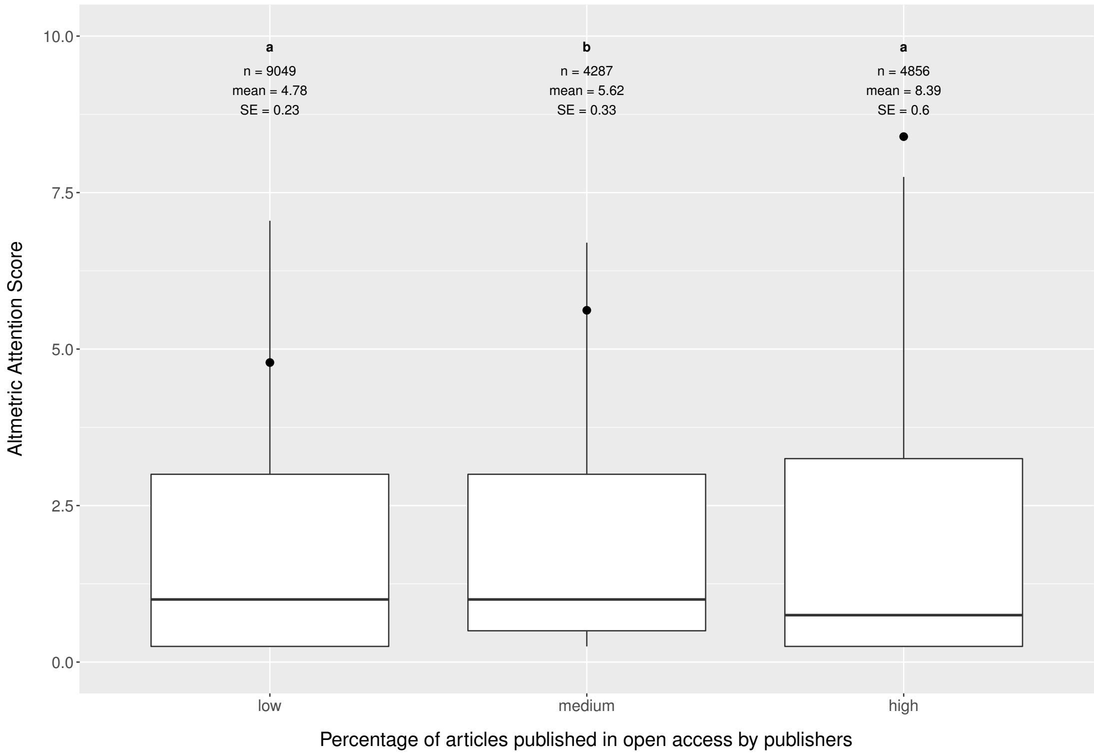

Figure 15: *Altmetric Attention Score* based on the percentage of open access content by publishers, based on the dataset retrieved from *Altmetric.com* (n = 18,192)

(publishers: n = 95, with low = 42, medium = 20, high = 33)

Means are displayed as black dots.

Listed categories are (as defined in subsection 3.6): 1) low (0-33.3% open access content), 2) medium (33.4-66.6%), 3) high (66.7-100%).

5. Conclusion and outlook

Altmetrics measure the societal impact of research in contrast to traditional metrics like citation counts. These new metrics are closely linked to social media as aggregators of altmetrics gather data from these as primary sources. Due to this focus on the public at large, new metrics might complement traditional metrics and provide additional information about impact of research outputs.

This study focused on altmetrics scores of articles published by German research institutions in the field of natural sciences in the years 2013 to 2017. Although results were statistically not significant, likely due to multiple outliers, they seemed to indicate that articles published in open access gain higher *Altmetric Attention Scores* compared to research outputs published in closed access, which became evident in higher mean values.

It was found that articles in gold open access had higher scores than those in hybrid, bronze open access or closed access, and that articles in green open access had higher *Altmetric Attention Score* than those in bronze open access and closed access. Over the years investigated, the *Altmetric Attention Scores* increased, which can be attributed to the emergence of social media and its increased usage by the public in general and also among scientists. The percentage of articles distributed in open access and the type of open access—gold, green, hybrid, bronze—depended strongly on publisher. Even when the majority of contents of two publishers was available under similar access status, in some cases huge differences in *Altmetric Attention Scores* have been determined—as was the case with *Oxford University Press* and the *American Physical Society*.

When taking a closer look at the publishers, percentage of content published in open access seemed to play an important role in the values gained in altmetrics scores. Publishers distributing a high percentage of their articles in open access also gained higher *Altmetric Attention Scores* for their articles than publishers with a medium or low percentage of open access content.

Altmetric Attention Scores differed between research areas, although the reasons behind this were not investigated. A detailed analysis into single disciplines was not possible despite the rather large sample encompassing a five-year span. Even so, the sample size per discipline was too small for thorough analysis due to the large number of disciplines.

The investigated measure *Altmetric Attention Score* was mostly composed of online mentions by tweeters, news outlets and to a smaller extent by Facebook accounts and blogs. It has to be kept in mind that the *Altmetric Attention Score* is a weighted measure and that news outlets and blog posts count more than tweeters and these in turn counts more than Facebook comments. Research was mostly mentioned by the public and by science communicators, and to a smaller extent by research scientists and practitioners. This was to be expected as the number of scientists in the population is comparatively small in terms of the whole population.

Overall, referring back to the primary research question, it seems that altmetrics scores are higher for open access content than for closed access content in the natural sciences for German research institutions.

One aspect that has to be kept in mind when conducting studies on altmetrics is that they are often criticized for their lack of integrity and data quality. This is a major concern as altmetrics data is gathered from a variety of primary sources. In recent years much has been improved concerning these issues. Nonetheless, replication of studies often is not possible. One reason for this is the dynamic nature of the World Wide Web; platforms can change over time or be replaced by new, more popular ones. Another reason is that posts or accounts on social media platforms can be deleted. Data providers should therefore further improve their methods for data collection and processing and ensure transparency of their algorithms and procedures.

Various research questions could be addressed more closely in the future:

- To gain further data, additional databases could be included into a wider study, both concerning the metadata—not just *Web of Science* but also *Scopus* could be included—as well as aggregators of altmetrics data. In this study, data was gathered from the primary aggregator *Altmetric.com* as this service offers data from various primary sources as well as a calculated composite indicator, the *Altmetric Attention Scores*, and background information such as which group of people—so-called cohorts—mentioned content. Another primary aggregator would be *Lagotto*, another large secondary one *Plum Analytics*. However, with including more sources of data, reliable and correct merging of this data would become an increasing challenge.

- Another similar study to the one conducted could be done in future years, then encompassing a larger amount of data including the years to come, thus possibly gaining new insights into altmetrics. The temporal scope of studies should not be chosen too widely as social media became popular only in recent years and open access also only evolved in the last decades. Moreover, a comparable study for the field of humanities would be conceivable, which should then include other document types such as books or contributions in anthologies. Another possibility would be the investigation of alternative research outputs like datasets and presentation slides.

- Further investigations based on single specific disciplines could be conducted, although samples would be very small for some fringe disciplines. Similarly, focused analysis of individual important journals in certain fields or mentions of research outputs on specific platforms could be conducted, both of which have already been investigated more closely in various studies. Writing a review article composing insights gained into altmetrics in various disciplines and in specific journals, in which these research outputs have been published, would also be conceivable.
- Another question, widely discussed and investigated, is the possible implementation of altmetrics into research evaluation as well as tenure and promotion. In some countries funding decisions are already not only based on traditional metrics like citation counts but also on altmetrics data of research outputs. These new metrics might complement the more traditional ones in providing additional information about the diverse impact research has had on the general public.

Overall, much research has already been published about altmetrics in recent years. However, many questions still remain to be investigated further and in more detail. Overall, altmetrics will certainly be an important research topic in years to come.

List of Figures

List of Tables

References

Adams, Jonathan (2012). *Collaborations: The rise of research networks*. In: Nature 490.7420, pp. 335–336. DOI: 10.1038/490335a.

Alhoori, Hamed; Choudhury, Sagnik Ray; Kanan, Tarek; Fox, Eedward A.; Furuta, Richard, & Giles, Lee C. (2015). *On the relationship between Open Access and altmetrics*. In: Proceedings of the iConference. URL: http://hdl.handle.net/2142/73451.

Alperin, Juan Pablo (2015). *Geographic variation in social media metrics: an analysis of Latin American journal articles*. In: Aslib Journal of Information Management 67.3, pp. 289–304. DOI: 10.1108/AJIM-12-2014-0176.

Alperin, Juan Pablo; Gomez, Charles J., & Haustein, Stefanie (2019). *Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles*. In: Public Understanding of Science 28.1, pp. 2–18. DOI: 10.1177/0963662518761733.

Altmetric.com (2015). *The donut and Altmetric Attention Score. An at-a-glance indicator of the volume and type of attention a research output has received*. Web Page. First published on 09/07/2015, last modified on 17/12/2018. URL: https://www.altmetric.com/about-our-data/the-donut-and-score/ (visited on 03/10/2019).

Archambault, Éric; Amyot, Didier; Deschamps, Philippe; Nicol, Aurore; Provencher, Françoise; Rebout, Lise, & Roberge, Guillaume (2014). *Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels1996–2013*. Science-Metrix, European Commission. URL: http://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf (visited on 03/10/2019).

Archambault, Éric; Campbell, David; Gingras, Yves, & Larivière, Vincent (2009). *Comparing bibliometric statistics obtained from the Web of Science and Scopus*. In: Journal of the American Society for Information Science and Technology 60.7, pp. 1320–1326. DOI: 10.1002/asi.21062.

Archambault, Éric; Côté, Grégoire; Struck, Brooke, & Voorons, Matthieu (2016). *Research impact of paywalled versus open access papers*. DigitalCommons@University of Nebraska - Lincoln. URL: http://digitalcommons.unl.edu/scholcom/29/ (visited on 03/10/2019).

Asemi, Asefeh & Heydari, Mahboobeh (2018). *Correlation between the Articles Citations in Web of Science (WoS) and the Readership Rate in Mendeley and Re-*

search Gate (RG). In: Journal of Scientometric Research 7.3, pp. 145–152. DOI: 10.5530/jscires.7.3.25.

Björk, Bo-Christer & Solomon, David (2012). *Open access versus subscription journals. A comparison of scientific impact*. In: BMC Medicine 10.73, pp. 1–10. DOI: 10.1186/1741-7015-10-73.

Black, Colin S.; Lehane, Daniel J.; Burns, Chris, & O'Donnell, Brian D. (2018). *An examination of the effect of open versus paywalled access publication on the disseminative impact and citation count of publications in intensive care medicine and anesthesia*. In: Journal of Critical Care 46, pp. 88–93. DOI: 10.1016/j.jcrc.2018.05.008.

Bornmann, Lutz (2013). *What is societal impact of research and how can it be assessed? a literature survey*. In: Journal of the American Society for Information Science 64.2, pp. 217–233. DOI: 10.1002/asi.22803.

– (2014). *Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics*. In: Journal of Informetrics 8.4, pp. 895–903. DOI: 10.1016/j.joi.2014.09.005.

– (2015a). *Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics*. In: Scientometrics 103.3, pp. 1123–1144. DOI: 10.1007/s11192-015-1565-y.

– (2015b). *Usefulness of altmetrics for measuring the broader impact of research. A case study using data from PLOS and F1000Prime*. In: Aslib Journal of Information Management 67.3, pp. 305–319. DOI: 10.1108/AJIM-09-2014-0115.

Bornmann, Lutz & Haunschild, Robin (2016). *To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics*. In: Online Information Review 40.4, pp. 529–543. DOI: 10.1108/OIR-09-2015-0314.

– (2018a). *Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data*. In: PLoS One 13.5. DOI: 10.1371/journal.pone.0197133.

– (2018b). *Alternative article-level metrics. The use of alternative metrics in research evaluation*. In: Embo Reports 19.12. DOI: 10.15252/embr.201847260.

Brewer, John (2011). *The impact of impact*. In: Research Evaluation 20.3, pp. 255–256. DOI: 10.3152/095820211X12941371876869.

Butler, Joseph S.; Kaye, I. David; Sebastian, Arjun S.; Wagner, Scott C.; Morrissey, Patrick B.; Schroeder, Gregory D.; Kepler, Christopher K., & Vaccaro, Alexander R. (2017). *The Evolution of Current Research Impact Metrics: From Bibliometrics* *to Altmetrics?* In: Clinical Spine Surgery 30.5, pp. 226–228. DOI: 10.1097/BSD.0000000000000531.

Costas, Rodrigo; Zahedi, Zohreh, & Wouters, Paul (2015). *Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective.* In: Journal of the Association for Information Science and Technology 66.10, pp. 2003–2019. DOI: 10.1002/asi.23309.

Creative Commons (2013). *Attribution 4.0 International (CC BY 4.0).* Web Page. First published on 25/11/2013. URL: https://creativecommons.org/licenses/by/4.0/ (visited on 03/10/2019).

Davis, Philip M. (2009). *Author-Choice Open-Access Publishing in the Biological and Medical Literature: A Citation Analysis.* In: Journal of the American Society for Information Science and Technology 60.1, pp. 3–8. DOI: 10.1002/asi.20965.

Davis, Philip M. & Fromerth, Michael J. (2006). *Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?* In: Scientometrics 71.2, pp. 203–215. DOI: 10.1007/s11192-007-1661-8.

Davis, Philip M.; Lewenstein, Bruce V.; Simon, Daniel H.; Booth, James G., & Connolly, Mathew J. L. (2008). *Open access publishing, article downloads, and citations: randomised controlled trial.* In: BMJ 337. DOI: 10.1136/bmj.a568.

DFG, (Deutsche Forschungsgemeinschaft) (2017). *DFG-Vordruck 12.20 – 02/17. Merkblatt Open Access Publizieren.* URL: http://www.dfg.de/formulare/12_20/index.jsp (visited on 03/10/2019).

– (2018). *DFG-Vordruck 2.00 – 11/18. Verwendungsrichtlinien - Allgemeine Bedingungen für Förderverträge mit der Deutschen Forschungsgemeinschaft e.V. (DFG).* URL: http://www.dfg.de/formulare/2_00/ (visited on 03/10/2019).

Di Girolamo, Nicola & Reynders, Reint Meursinge (2017). *Health care articles with simple and declarative titles were more likely to be in the Altmetric Top 100.* In: Journal of Clinical Epidemiology 85, pp. 32–36. DOI: 10.1016/j.jclinepi.2016.11.018.

DORA (2013). *San Francisco Declaration on Research Assessment (DORA).* Web Page. Declaration was first published in May 2013, launching of the new web page was announced 07/02/2018. URL: https://sfdora.org/read/ (visited on 03/10/2019).

Elmore, Susan A. (2018). *The Altmetric Attention Score: What Does It Mean and Why Should I Care?* In: Toxicologic Pathology 46.3, pp. 252–255. DOI: 10.1177/0192623318758294.

Erdt, Mojisola; Nagarajan, Aarthy; Sin, Sei-Ching Joanna, & Theng, Yin-Leng (2016). *Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media*. In: Scientometrics 109.2, pp. 1117–66. DOI: 10.1007/s11192-016-2077-0.

European Commission (2018). *H2020 Policy Support Facility, Mutual Learning Exercise on Open Science: Altmetrics and Rewards - Thematic report on Different types of Altmetrics*. URL: https://rio.jrc.ec.europa.eu/en/policy-support-facility/mle-open-science-altmetrics-and-rewards (visited on 03/10/2019).

Eysenbach, Gunther (2006). *Citation advantage of open access articles*. In: PLoS Biology 4.5. e157. DOI: 10.1371/journal.pbio.0040157.

Fuchs, Christian & Sandoval, Marisol (2013). *The Diamond Model of Open Access Publishing: Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-Commercial, Non-Profit Open Access Serious*. In: tripleC 11.2. DOI: 10.31269/triplec.v11i2.502.

Garnier, Simon (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.0. URL: https://CRAN.R-project.org/package=viridis (visited on 03/10/2019).

Gauch, Stephan & Blümel, Clemens (2016). *OPENing UP new methods, indicators and tools for peer review, impact measurement and dissemination of research results*. European Commission, OpenUP. DOI: 10.5281/zenodo.1146618.

Gordon, Gregg; Lin, Jennifer; Cave, Richard, & Dandrea, Ralph (2015). *The question of data integrity in article-level metrics*. In: PLoS Biology 13.8. e1002161. DOI: 10.1371/journal.pbio.1002161.

Gross, Juergen & Ligges, Uwe (2015). *nortest: Tests for Normality*. R package version 1.0-4. URL: https://CRAN.R-project.org/package=nortest (visited on 03/10/2019).

Gumpenberger, Christian; Glänzel, Wolfgang, & Gorraiz, Juan (2016). *The ecstasy and the agony of the altmetric score*. In: Scientometrics 108.2, pp. 977–982. DOI: 10.1007/s11192-016-1991-5.

Haustein, Stefanie (2016). *Grand challenges in altmetrics: heterogeneity, data quality and dependencies*. In: Scientometrics 108.1, pp. 413–423. DOI: 10.1007/s11192-016-1910-9.

Haustein, Stefanie; Bowman, Timothy D., & Costas, Rodrigo (2015a). *Interpreting "altmetrics": viewing acts on social media through the lens of citation and social theories*. In: Theories of Informetrics: A Festschrift in Honor of Blaise Cronin. Ed. by Cassidy R. Sugimoto. arXiv: 1502.05701.

Haustein, Stefanie; Costas, Rodrigo, & Larivière, Vincent (2015b). *Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns*. In: PloS One 10.3. e0120495. DOI: 10.1371/journal.pone.0120495.

Haustein, Stefanie & Larivière, Vincent (2015). *The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects*. In: Incentives and Performance: Governance of Research Organizations. Ed. by Isabell M. Welpe; Jutta Wollersheim; Stefanie Ringelhan, & Margit Osterloh. Cham: Springer International Publishing, pp. 121–139. DOI: 10.1007/978-3-319-09785-5_8.

Hicks, Diana; Wouters, Paul; Waltman, Ludo; Rijcke, Sarah de, & Rafols, Ismael (2015). *Bibliometrics: the Leiden Manifesto for Research Metrics*. In: Nature 520.7548. URL: http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-researchmetrics-1.17351 (visited on 03/10/2019).

Holmberg, Kim (2015). *Altmetrics for Information Professionals : Past, Present and Future*. Witney, Oxford, UK: Chandos Publishing. DOI: 10.1016/C2014-0-03379-5.

Holmberg, Kim & Vainio, Julia (2018). *Why do some research articles receive more online attention and higher altmetrics? Reasons for online success according to the authors*. In: Scientometrics 116.1, pp. 435–447. DOI: 10.1007/s11192-018-2710-1.

Huang, Wenya; Wang, Peiling, & Wu, Qiang (2018). *A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals*. In: PLoS One 13.4. e0194962. DOI: 10.1371/journal.pone.0194962.

Kay, Matthew & Wobbrock, Jacob O. (2019). *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs*. R package version 0.10.6. DOI: 10.5281/zenodo.594511.

Kurtz, Michael J. & Bollen, Johan (2010). *Usage bibliometrics*. In: Annual Review of Information Science and Technology 44.1, pp. 1–64. DOI: 10.1002/aris.2010.1440440108.

Kurtz, Michael J.; Eichhorn, Guenther; Accomazzi, Alberto; Grant, Carolyn; Demleitner, Markus; Henneken, Edwin, & Murray, Stephan S. (2005). *The effect of use and access on citations*. In: Information Processing and Management 41, pp. 1395–1402. DOI: 10.1016/j.ipm.2005.03.010.

Lin, Jennifer & Fenner, Martin (2013). *The many faces of article-level metrics*. In: Bulletin of the Association for Information Science and Technology 39.4, pp. 27–30. DOI: 10.1002/bult.2013.1720390409.

Littler, Katherina; Boon, Wee-Ming; Carson, Gail; Deportere, Evelyn; Mathewson, Sophie; Mietchen, Daniel; Moorthy, Vasee S.; O'Connor, Denise; Roth, Cathy, & Segovia, Carlos (2017). *Progress in promoting data sharing in public health emergencies*. In: Bulletin of the World Health Organization 95.4, p. 243. DOI: 10.2471/BLT.17.192096.

Moed, Henk F. & Halevi, Gali (2016). *On full text download and citation distributions in scientific-scholarly journals*. In: Journal of the Association for Information Science and Technology 67.2, pp. 412–431. DOI: 10.1002/asi.23405.

Mukherjee, Bhaskar; Subotić, Siniša, & Chaubey, Ajay Kumar (2018). *And now for something completely different: the congruence of the Altmetric Attention Score's structure between different article groups*. In: Scientometrics 114.1, pp. 253–275. DOI: 10.1007/s11192-017-2559-8.

Nicholas, David; Huntington, Paul; Dobrowolski, Tom; Rowlands, Ian; Hamid, R. Jamali M., & Polydoratou, Panayiota (2005). *Revisiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication*. In: Information Processing and Management 41.6, pp. 1441–1461. DOI: 10.1016/j.ipm.2005.03.014.

NISO, (National Information Standards Organization) (2016). *NISO RP-25-2016 Outputs of the NISO Alternative Assessment Metrics Project*. URL: https://www.niso.org/publications/rp-25-2016-altmetrics (visited on 03/10/2019).

Norris, Michael; Oppenheim, Charles, & Rowland, Fytton (2008). *The citation advantage of open-access articles*. In: Journal of the American Society for Information Science and Technology 59.12, pp. 1963–1972. DOI: 10.1002/asi.20898.

Ortega, José Luis (2017). *Reliability and accuracy of altmetric providers: a comparison among Altmetric, PlumX and Crossref Event Data*. In: Scientometrics 116.3, pp. 2123–2138. DOI: 10.1007/s11192-018-2838-z.

Pacheco, André; Sousa, Alexandre; Yanai, Angela; Lopes, Susana, & Machado, Luís (2018). *Metrics and altmetrics: An exploratory study of a possible correlation between the most cited papers in open and restricted access in 2016-2018*. In: ACM International Conference Proceeding Series, pp. 258–264. DOI: 10.1145/3284179.3284224.

Pinfield, Stephen (2015). *Making Open Access work: The "state-of-the-art" in providing Open Access to scholarly literature*. In: Online Information Review 39.5, pp. 604–636. DOI: 10.1108/OIR-05-2015-0167.

Pinfield, Stephen; Salter, Jennifer, & Bath, Peter A. (2016). *The "total cost of publication" in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions*. In: Journal of the Association for Information Science and Technology 67.7, pp. 1751–1766. DOI: 10.1002/asi.23446.

Piwowar, Heather; Priem, Jason; Larivière, Vincent; Alperin, Juan Pablo; Matthias, Lisa; Norlander, Bree; Farley, Ashley; West, Jevin, & Haustein, Stefanie (2018). *The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles*. In: PeerJ 6. DOI: 10.7717/peerj.4375.

Priem, Jason; Taraborelli, Dario; Groth, Paul, & Neylon, Cameron (2010). *Altmetrics: a manifesto*. Web Page. First published on 26/10/2010, last modified on 28/09/2011. URL: http://altmetrics.org/manifesto/ (visited on 03/10/2019).

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/ (visited on 03/10/2019).

REF, (Research Excellence Framework) (2011). *Assessment Framework and Guidance on Submissions. REF 02.2011*. URL: https://www.ref.ac.uk/2014/pubs/2011-02/ (visited on 03/10/2019).

– (2014). *REF 2014 - Key facts leaflet*. URL: https://www.ref.ac.uk/2014/pubs/keyfacts/ (visited on 03/10/2019).

Robinson-García, Nicolás; Arroyo-Machado, Wenceslao; Moed, Henk F., & Torres-Salinas, Daniel (2018a). *Do altmetrics promote Open Access? An exploratory analysis on altmetric differences between types of access in the field of Physics*. DOI: 10.5281/zenodo.1414022.

Robinson-García, Nicolás; Leeuwen, Thed N. van, & Ràfols, Ismael (2018b). *Using altmetrics for contextualised mapping of societal impact: From hits to networks*. In: Science and Public Policy 45.6, pp. 815–826. DOI: 10.1093/scipol/scy024.

Rowlands, Ian; Nicholas, David; Russell, Bill; Canty, Nicholas, & Watkinson, Anthony (2011). *Social media use in the research workflow*. In: Learned Publishing 24.3, pp. 183–195. DOI: 10.1087/20110306.

RStudio Team (2016). *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA. URL: http://www.rstudio.com/ (visited on 03/10/2019).

Samuel, Gabrielle N. & Derrick, Gemma E. (2015). *Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014*. In: Research Evaluation 24.3, pp. 229–241. DOI: 10.1093/reseval/rvv007.

Schimmer, Ralf; Geschuhn, Kai Karin, & Vogler, Andreas (2015). *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*. DOI: 10.17617/1.3.

Sugimoto, Cassidy R.; Work, Sam; Larivière, Vincent, & Haustein, Stefanie (2017). *Scholarly use of social media and altmetrics: a review of the literature*. In: Journal of the American Society for Information Science and Technology 68.9, pp. 2037–2062. DOI: 10.1002/asi.23833.

Swan, Alma (2010). *Open access citation advantage. Studies and results to date*. URL: https://eprints.soton.ac.uk/id/eprint/268516 (visited on 03/10/2019).

Taylor, Mike (2013a). *Exploring the Boundaries: How Altmetrics Can Expand Our Vision of Scholarly Communication and Social Impact*. In: Information Standards Quarterly 25 (2), pp. 27–32. URL: https://www.niso.org/niso-io/2013/06/exploring-boundaries (visited on 03/10/2019).

– (2013b). *Towards a common model of citation: Some thoughts on merging altmetrics and bibliometrics*. In: Research Trends 35, pp. 19–22. URL: https://www.researchtrends.com/issue-35-december-2013/towards-a-common-model-of-citation-some-thoughts-on-merging-altmetrics-and-bibliometrics/ (visited on 03/10/2019).

Thelwall, Mike (2018). *Early Mendeley readers correlate with later citation counts*. In: Scientometrics 115.3, pp. 1231–1240. DOI: 10.1007/s11192-018-2715-9.

Thelwall, Mike & Kousha, Kayvan (2015). *Web indicators for research evaluation. Part 2: Social media metrics*. In: El Profesional de La Información 24.5. DOI: 10.3145/epi.2015.sep.09.

Thelwall, Mike; Kousha, Kayvan; Dinsmore, Adam, & Dolby, Kevin (2016). *Alternative metric indicators for funding scheme evaluations*. In: Aslib Journal of Information Management 68.1, pp. 2–18. DOI: 10.1108/ajim-09-2015-0146.

Tunger, Dirk (2017). *Sichtbarkeit und Wahrnehmung von Open-Access-Veröffentlichungen unter bibliometrischen Aspekten*. In: Praxishandbuch Open Access. Ed. by Konstanze Söllner & Bernhard Mittermaier. Berlin; Boston: De Gruyter Saur. DOI: 10.1515/9783110494068-005.

Van Noorden, Richard (2014). *Online collaboration: Scientists and the social network*. In: Nature 512.7513, pp. 126–129. DOI: 10.1038/512126a.

Wang, Xianwen; Liu, Chen; Mao, Wenli, & Fang, Zhichao (2015). *The open access advantage considering citation, article usage and social media attention*. In: Scientometrics 103.2, pp. 555–564. DOI: 10.1007/s11192-015-1547-0.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: http://ggplot2.org (visited on 03/10/2019).

– (2018). *scales: Scale Functions for Visualization*. R package version 1.0.0. URL: https://CRAN.R-project.org/package=scales (visited on 03/10/2019).

Wilsdon, James; Allen, Liz; Belfiore, Eleonora; Campbell, Philip; Curry, Stephen, & Hill, Steven et al. (2015). *The Metric Tide*. Report of the Independent Review of the Role of Metrics in Research Assessment and Management: HEFCE. DOI: 10.13140/RG.2.1.4929.1363.

Wilsdon, James; Bar-Ilan, Judit; Frodeman, Robert; Lex, Elisabeth; Peters, Isabella, & Wouters, Paul (2017). *Next-generation metrics. responsible metrics and evaluation for Open Science*. Report of the European Commission Expert Group on Altmetrics. European Commission. Brüssel. URL: http://hdl.handle.net/1887/58317.

Wobbrock, Jacob O.; Findlater, Leah; Gergle, Darren, & Higgins, James J. (2011). *The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures*. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). New York: ACM Press, pp. 143–146. URL: http://depts.washington.edu/aimgroup/proj/art/ (visited on 03/10/2019).

Wooldridge, Jenny & King, Mike B. (2018). *Altmetric scores: An early indicator of research impact*. In: Journal of the Association for Information Science and Technology. DOI: 10.1002/asi.24122.

Wren, Jonathan D. (2005). *Open access and openly accessible: A study of scientific publications shared via the internet*. In: BMJ 330, p. 1128. DOI: 10.1136/bmj.38422.611736.E0.

Zahedi, Zohreh & Costas, Rodrigo (2018). *General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators*. In: PLoS One 13.5. DOI: 10.1371/journal.pone.0197326.

Zahedi, Zohreh; Costas, Rodrigo, & Wouters, Paul (2017). *Mendeley readership as a filtering tool to identify highly cited publications*. In: Journal of the Association for Information Science and Technology 68.10, pp. 2511–2521. DOI: 10.1002/asi.23883.

A. Appendix

The following abbreviations were used in the appendix:

- Alt: *Altmetric.com*
- WoS: *Web of Science*

Figure 2

Percentage of articles per publication year

publication year	percentage WoS	percentage Alt
2013	17.2	13.2
2014	18.8	16.7
2015	19.9	19.8
2016	21.8	24.0
2017	22.4	26.4

Figure 3

Percentage of articles per access status

publication year	access status	percentage WoS	percentage Alt
2013	gold	3.1	4.6
	green	0.5	0.5
	hybrid	1.4	2.4
	bronze	15.5	26.0
	closed	78.9	66.3
	unclear	0.6	0.2
2014	gold	3.9	6.7
	green	0.4	0.5
	hybrid	2.9	4.7
	bronze	15.7	34.2
	closed	76.2	53.7
	unclear	1.0	0.1
2015	gold	5.5	7.7
	green	0.6	0.8
	hybrid	4.5	6.1

Percentage of articles per access status (continued)

publication year	access status	percentage WoS	percentage Alt
2015	bronze	16.6	33.4
	closed	72.1	51.9
	unclear	0.7	0.1
2016	gold	7.0	10.9
	green	0.7	1.1
	hybrid	4.6	5.7
	bronze	17.2	32.2
	closed	69.9	50.1
	unclear	0.6	0.0
2017	gold	8.3	14.4
	green	0.7	1.3
	hybrid	4.2	5.4
	bronze	17.5	32.1
	closed	68.5	46.8
	unclear	0.8	0.1

Figure 4

Percentage of articles per research area

research area	percentage WoS	percentage Alt
Biochemistry & Molecular Biology	2.0	5.1
Biomedicine	4.6	5.8
Chemistry	15.8	19.6
Computer Science	12.8	14.2
Energy & Fuels	7.2	5.5
Engineering	35.0	24.3
Environmental Sciences & Ecology	4.0	6.2
Life Sciences	4.8	8.0
Life Sciences & Biomedicine - Other Topics	3.3	8.4
Materials Science	38.7	41.0
Physical Sciences	16.3	14.8
Physics	22.7	30.9
Technology	33.1	30.9

Figure 5

Percentage of articles per research area and access status

research area	access status	percentage WoS	percentage Alt
Biochemistry & Molecular Biology	gold	7.1	9.0
	green	1.0	1.2
	hybrid	3.0	3.0
	bronze	34.7	42.5
	closed	54.1	44.3
	unclear	0.1	0.0
Biomedicine	gold	2.2	4.5
	green	0.7	1.5
	hybrid	3.6	5.5
	bronze	13.4	14.5
	closed	78.4	73.7
	unclear	1.7	0.4
Chemistry	gold	4.2	6.0
	green	0.5	0.6
	hybrid	6.5	7.4
	bronze	11.7	18.0
	closed	76.8	68.1
	unclear	0.4	0.0
Computer Science	gold	2.0	4.1
	green	0.9	1.2
	hybrid	2.6	3.4
	bronze	27.9	42.1
	closed	65.4	58.9
	unclear	1.2	0.3
Energy & Fuels	gold	5.6	11.7
	green	0.7	1.2
	hybrid	5.4	9.2
	bronze	7.2	11.5
	closed	80.7	66.4
	unclear	0.4	0.0
Engineering	gold	2.4	3.3
	green	0.6	1.0
	hybrid	2.1	3.9

Percentage of articles per research area and access status (continued)

research area	access status	percentage WoS	percentage Alt
Engineering	bronze	8.7	14.8
	closed	85.5	76.9
	unclear	0.7	0.1
Environmental Sciences & Ecology	gold	0.5	0.5
	green	1.2	1.8
	hybrid	6.8	10.1
	bronze	19.6	29.8
	closed	71.5	57.8
	unclear	0.4	0.0
Life Sciences	gold	1.4	1.9
	green	0.9	1.5
	hybrid	6.2	7.7
	bronze	25.3	43.9
	closed	65.7	45.0
	unclear	0.6	0.1
Life Sciences & Biomedicine - Other Topics	gold	28.9	36.9
	green	1.2	1.6
	hybrid	7.0	7.4
	bronze	29.5	32.5
	closed	32.8	21.5
	unclear	0.6	0.1
Materials Science	gold	4.6	5.4
	green	0.5	0.8
	hybrid	4.7	5.3
	bronze	18.5	39.9
	closed	71.0	48.7
	unclear	0.7	0.0
Physical Sciences	gold	4.8	8.6
	green	0.7	0.7
	hybrid	4.9	8.1
	bronze	17.4	31.6
	closed	71.7	50.9
	unclear	0.5	0.0
Physics	gold	4.0	2.8

Percentage of articles per research area and access status (continued)

research area	access status	percentage WoS	percentage Alt
Physics	green	0.5	0.7
	hybrid	4.6	5.4
	bronze	29.5	55.3
	closed	61.2	35.8
	unclear	0.3	0.0
Technology	gold	8.1	11.3
	green	0.5	0.8
	hybrid	3.9	5.6
	bronze	13.7	23.9
	closed	73.3	58.2
	unclear	0.5	0.1

Figure 6

Percentage of articles distributed by major publishers

publisher	percentage WoS	percentage Alt
ACS (American Chemical Society)	7.5	11.3
APS (American Physical Society)	5.5	12.3
Elsevier	29.1	18.5
IEEE (Institute of Electrical and Electronics Engineers)	6.9	4.7
IOP (Institute of Physics)	2.7	3.1
MDPI (Multidisciplinary Digital Publishing Institute)	2.1	3.6
Oxford University Press	0.8	2.3
RSC (Royal Society of Chemistry)	2.9	3.5
Springer	10.1	6.8
Taylor & Francis	2.1	1.3
Wiley	10.5	14.9

Figure 7

Percentage of articles distributed by major publishers in different access status

publisher	access status	percentage WoS	percentage Alt
ACS (American Chemical Society)	gold	0.0	0.0
	green	0.5	0.9
	hybrid	1.6	1.4
	bronze	13.4	20.1
	closed	83.8	77.6
	unclear	0.7	0.0
APS (American Physical Society)	gold	0.0	0.0
	green	0.9	0.8
	hybrid	1.0	0.9
	bronze	70.3	90.8
	closed	27.8	7.5
	unclear	0.0	0.0
Elsevier	gold	0.8	0.6
	green	0.7	1.3
	hybrid	3.1	5.8
	bronze	8.9	19.7
	closed	86.1	72.7
	unclear	0.4	0.0
IEEE (Institute of Electrical and Electronics Engineers)	gold	0.0	0.0
	green	0.7	1.9
	hybrid	0.0	0.1
	bronze	19.9	38.8
	closed	79.2	59.3
	unclear	0.2	0.0
IOP (Institute of Physics)	gold	0.2	0.0
	green	0.4	0.9
	hybrid	9.4	13.2
	bronze	17.7	38.2
	closed	72.0	47.7
	unclear	0.3	0.2

Percentage of articles distributed by major publishers in different access status (continued)

publisher	access status	percentage WoS	percentage Alt
MDPI (Multidisciplinary Digital Publishing Institute)	gold	100.0	100.0
Oxford University Press	gold	0.0	0.0
	green	0.9	1.2
	hybrid	1.5	1.5
	bronze	76.8	87.9
	closed	20.7	9.4
	unclear	0.2	0.0
RSC (Royal Society of Chemistry)	gold	0.0	0.0
	green	0.5	0.3
	hybrid	31.4	33.7
	bronze	17.4	21.0
	closed	50.7	45.0
	unclear	0.1	0.0
Springer	gold	2.9	6.0
	green	0.4	0.6
	hybrid	4.7	8.5
	bronze	13.6	26.7
	closed	78.2	58.2
	unclear	0.1	0.0
Taylor & Francis	gold	0.5	0.0
	green	0.9	1.2
	hybrid	0.0	0.0
	bronze	9.6	19.1
	closed	88.7	79.7
	unclear	0.4	0.0
Wiley	gold	0.7	0.7
	green	0.5	0.6
	hybrid	2.5	4.1
	bronze	9.5	11.8
	closed	86.8	82.9

Percentage of articles distributed by major publishers in different access status (continued)

publisher	access status	percentage WoS	percentage Alt
Wiley	unclear	0.0	0.0

Figure 8

Percentage of total posts by data source

data source	percentage
Delicious users	0.0
Facebook accounts	4.5
blogs	2.0
internet forum users	0.0
Google+ users	0.9
LinkedIn users	0.0
news outlets	8.0
peer review sites	0.1
Pinterest users	0.0
policy sources	0.3
questions, answers or comments on Stack Exchange sites	0.0
Reddit users	0.1
research highlight platforms	0.1
Twitter users	66.8
YouTube channels	0.1
Sina Weibo	0.1
Wikipedia (pages on Wikipedia)	0.7

Figure 9

Percentage of total cohorts by cohort

cohort	percentage
science communicators	26.8
the public	66.3
practitioners	1.7
research scientists	5.2

Figure 10

Altmetric Attention Score per publication year

year	letter	n	mean	SE	1st quartile	median	3rd quartile
2013	a	2411	5.50	0.37	0.50	1.3	4.0
2014	bc	3048	4.96	0.31	0.25	1.0	3.0
2015	b	3610	4.72	0.24	0.25	1.0	3.0
2016	bc	4378	6.70	0.43	0.25	1.0	3.0
2017	c	4809	6.99	0.61	0.50	1.0	2.5

abbreviations: n (sample size), SE (standard error)

Figure 11

Altmetric Attention Score for open and closed access

access status	letter	n	mean	SE	1st quartile	median	3rd quartile
open	a	8695	7.08	0.37	0.50	1.0	3.0
closed	a	9561	4.90	0.22	0.25	1.0	3.0

abbreviations: n (sample size), SE (standard error)

Figure 12

Altmetric Attention Score per access status

access status	letter	n	mean	SE	1st quartile	median	3rd quartile
gold	a	1764	3.31	1.53	0.5	1.5	6.3
green	ab	166	11.45	3.01	0.5	1.5	5.7
hybrid	b	931	8.33	0.72	0.5	1.4	5.1
bronze	c	5834	4.87	0.25	0.3	0.8	2.3
closed	d	9561	4.90	0.22	0.3	1.0	3.0

abbreviations: n (sample size), SE (standard error)

Figure 13

Altmetric Attention Score per research area

research area	letter	n	mean	SE	1st quartile	median	3rd quartile
Biochem	a	934	10.93	2.24	0.5	1.6	7.0
Biomed	bh	1060	3.92	0.55	0.5	1.0	3.0
Chem	c	3573	7.27	0.39	0.5	1.0	3.6
Comp	c	2578	3.94	0.29	0.5	1.0	3.0
Energy	a	1013	6.65	0.59	0.5	1.9	5.0
Eng	d	4436	4.07	0.33	0.3	1.0	3.0
Env	e	1140	14.85	1.33	1.0	3.0	10.2
LS	f	1460	9.69	0.60	0.5	2.0	8.0
LSBO	a	1537	22.88	1.86	1.0	4.6	18.5
Mat	d	7500	5.04	0.21	0.3	0.8	3.0
Physical	bc	2700	4.39	0.29	0.5	1.0	3.0
Physics	g	5635	4.22	0.23	0.3	0.5	1.8
Tech	h	5647	5.29	0.28	0.5	1.0	3.0

abbreviations: n (sample size), SE (standard error), Biochem (Biochemistry & Molecular Biology), Biomed (Biomedicine), Chem (Chemistry), Compu (Computer Science), Engergy (Energy & Fuels), Eng (Engineering), Env (Environmental Sciences & Ecology), LS (Life Sciences), LSBO (Life Sciences & Biomedicine - Other Topics), Mat (Materials Science), Physical (Physical Sciences), Tech (Technology)

Figure 14

Altmetric Attention Score based on major publishers

publisher	letter	n	mean	SE	1st quartile	median	3rd quartile
ACS	a	2067	7.38	0.78	0.5	1.0	3.0
APS	b	2255	1.43	0.13	0.3	0.5	0.8
Elsevier	ac	3378	3.24	0.17	0.5	1.0	3.0
IEEE	d	859	4.61	0.67	0.5	1.9	3.0
IOP	ce	568	3.94	0.53	0.3	0.8	3.0
MDPI	ac	664	2.44	0.33	0.5	1.0	1.9
OUP	f	413	8.54	0.86	1.3	3.8	10.8
RSC	ac	647	4.64	0.55	0.5	1.0	3.0
Springer	a	1238	3.63	0.43	0.5	1.0	3.0
TF	ace	241	3.73	1.05	0.5	1.0	1.9
Wiley	e	2725	4.82	0.29	0.3	0.5	3.1

abbreviations: n (sample size), SE (standard error), ACS (American Chemical Society), APS (American Physical Society), IEEE (Institute of Electrical and Electronics Engineers), IOP (Institute of Physics), MDPI (Multidisciplinary Digital Publishing Institute), OUP (Oxford University Press), RSC (Royal Society of Chemistry), TF (Taylor & Franics)

Figure 15

Altmetric Attention Score based on the percentage of open access content by publishers

category	letter	n	mean	SE	1st quartile	median	3rd quartile
low	a	9049	4.78	0.23	0.3	1.0	3.0
medium	b	4287	5.62	0.33	0.5	1.0	3.0
high	a	4856	8.39	0.60	0.3	0.8	3.3

abbreviations: n (sample size), SE (standard error); categories: low percentage of open access content (0-33.3%), medium percentage (33.4-66.6%), high percentage (66.7-100%)