

# How open is “open”? A bibliometric study on the literature of Open Science and Open Access

Giannis Tsakonas<sup>1</sup>, Sergios Lenis<sup>2</sup> and Moses Boudourides<sup>3,4</sup>

<sup>1</sup> Library & Information Center, University of Patras, Greece

<sup>2</sup> University of Patras, Greece

<sup>3</sup> Science Division, New York University Abu Dhabi, U.A.E.

<sup>4</sup> School of Professional Studies, Northwestern University, USA

**Abstract:** After twenty years of research in the field of Open Access and Open Science, there is a critical part of literature, but its characteristics have not been thoroughly explored. In this study we explore the Open Access and Open Science literature, as published the period 1999 - 2018 and indexed in the bibliographic database of Web of Science. The total amount of publications that we explore is 2.846 and our analyses run on two levels: first, in a macroscopic point of view, we explore statistic properties of the major entities of this field, namely publications, authors and sources, and how these have been developed over these two decades. We mainly focus on the access status of these entities, as well as we explore if there is evidence of citation advantage of openly accessible publications. Second, in the microscopic point of view, we check the temporal characteristics of certain groups of authors and sources to see which publication practices have appeared this period. Our study shows that the field is expanding, not just in absolute publication numbers, but also in terms of authors express their interest. However, one of the main findings of this study is that authors that publish on these topics keep limited connection.

**Keywords:** Open Access, Open Science, bibliometric study, citation advantage, openness

---

## 1. Introduction

The field of Open Access has stirred a heated debate in the scholarly communication system about the means and the processes of accessing scientific information. Many aspects of this debate deal with the democratization of research, the upgrade of scientific and research performance and the reasonable and ethical use of resources. In a paper that summarizes the impact of OA, Tennant et al. [1] underline the inequality and fragility of the system, claiming that “Current levels of access in the developing world are insufficient and unstable, and OA has the potential to foster the development of stable research ecosystems.” Access, however,

to the products of research performing entities cannot be separated by the state of production and processes. As a natural consequence, the scholarly communication world realized that in order to make changes in OA, radical changes to the foundations of the system are required, including research assessment processes and criteria, new rewards, skills, and so on. This has formed the notion of Open Science. Lately, various political moves and decisions, either from research funding organizations or from institutions alone, have shifted the discourse on the quality of OA that the system wants. The various potential roads of OA, together with the policy and the infrastructure prerequisites they carry together, are shaping future scholarly communication. Therefore, the need for opinion exchange and informed audiences is imperative and this need is expressed as a growing part of the literature that discusses the aspects of OA and OS.

Many OS advocates are calling for a change of culture that will truly enact its principles. As long as cultural traits of the existing paradigm are in force, OS is impeded. In order to exemplify the benefits of OS, scholarly communication actors will need to commit to its principles and adopt its practices consistently and robustly [2, 3]. However, several contradictions are being recorded. Mentions about articles that discuss OA developments published behind the tolls of platforms of emphatically criticized commercial publishers are feeding the social media timelines resulting in comments of wry nature [4, 5]. Apart from the ethical viewpoints, it is important that every part of the discourse for OS remains open and accessible to anyone who wishes to contribute.

The topics of this debate are highly complex and require thorough examination. However, after twenty years of research, there is a critical part of the literature and its characteristics have not been explored. In this study, we perform a quantitative analysis of 2,846 publications that focus on Open Science and Open Access, as indexed in Web of Science, to explore the development of this literature over the last twenty years. We question how much of this literature is open and how key entities have been developed over these years. We are interested to see the variations per access type, as well as to explore whether other questions of the topic, such as the citation advantage claim, can be answered for this field as well. This paper is structured as this: Section 2 presents the background of the study by highlighting several studies on the growth of OA/OS field. Section 3 presents the research setting, as well as delineates the main research questions. Section 4 presents the findings of the study in two large subsections. Section 5 is the area for discussing these findings, while in the last section, we conclude our study.

## 2. Background

Open Access and Open Science are considered as movements for the restoration of scholarly communication values and the establishment of links with society and innovation. Research funding organizations have pushed towards the adoption of progressive policies, including the PlanS implementation and Wellcome Trust policies. OS is often greeted as a beneficial set of practices for researchers, as McKiernan et al. [6] have supported. They delineate the benefits of OS in an ecological system of openness, from Open Access to Open Data and Open Source, and conclude that these can take the form of increased attention to their work, higher impact, more collaboration opportunities, etc. The Open Science Monitor of European Commission [7] monitors the progress on these areas, especially on the growth of OA, the management of research data and the public engagement in scientific processes. At the same time, SPARC Europe presents regular reviews of the European OS policy landscape to understand how the national dynamics are implemented. In its most recent version [8], the review concludes that European legislation on the use of public sector data has been a driver of developments in research data adoption, while the national policies in place have been increased.

The ability to search a critical mass in almost every scientific field that someone can analyze and draw conclusions on, as well as the proliferation of data gathering or analytic tools, has led to numerous bibliometric studies which explore the growth of OA publications in disciplinary or national contexts. Growth and understanding of the situation is the prime interest of the researchers of this field, according to a study by Rodrigues, Taga and Passos [9]. In their content analysis study, the studies of the concepts of Growth and Perception represent nearly 40% of their data, followed by papers that are concerned about the Economics and the Impact of the field.

Björk et al. [10] have recorded one of the first percentages of OA availability. In 2009 they reported an approximately 20% of accessible scientific literature with a breakdown of 8% for formal and 12% for availability in repositories and web services. Laakso et al. [11] have found that the OA journals and their publications grow at much faster rates than the rest of the literature and have stabilized their production, availability and exposure. Archambault et al. [12] have provided a thorough, large scale review of the OA availability, which reaches up to 50%. The study attributes such high rates partially to the backfiling of publications in repositories and websites and to the growth of Gold OA publications. In a recent study, Piwowar et al. [13] have projected the growth of OA and predict that in 2025 70% of the readings of the literature will be in OA and that the rate of these publications will climb from

30% to 45%. The current 30% has been confirmed by a previous study of Piwowar et al. [14] who analyzed 300.000 records from three distinct sources to find that this percentage of scientific publications is openly accessible, even with low percentages of Green and relatively high of Bronze OA status. The authors of this study believe that the gradient increase of opening publications will have collateral effects on library subscriptions and big deal negotiations.

Similar rates of availability have been recorded in Sweden by Fathli, Lundén and Sjögarde [15]. In their study it was found that only 25% of Swedish publications were openly available. The authors noted that the varied practices of many authors create gray zones between the various access types. According to them, this results to conceptual ambiguity that stalls a unified comprehension of OA. Rovira, Urbano and Abadal [16] examined the OA publications in Catalonia and reported an impressive 70% availability of at least one version for each publication. This percentage relies heavily on the availability of publication in academic social media services, such as ResearchGate and Academia.edu, and subject repositories. The authors however have noted that there is margin for further compliance with the Spanish Science Act, that requires, as other national legislation, immediate public availability.

Breugelmans et al. [17] estimate that there is a 70% availability of medical papers on the web. The authors also support that there is a significant citation advantage for OA publications compared to paywalled in this field, while Kurata et al. [18] have shown that the rates of OA publications in the biomedical field have been increased significantly from 2006 to 2010. Liu and Li [19] present evidence of rapid growth of OA publications in social sciences, especially after 2009. Gargouri et al. [20] estimated that the rate of openly accessible publications in 14 disciplines is -on average- 25%, with the highest rates corresponding to Mathematics, Social Science and Earth & Space Sciences. The Open Science Monitor [7], using data from Scopus and Unpaywall, has recorded high rates for medical sciences, mathematics and agricultural sciences. Morison [21] has exhibited data of “dramatic growth” in 2018. She has analyzed data from various prominent web services and repositories and among her findings is the advancement of biomedical preprint servers and the increase of SCOAP3, a consortium-based funding project in the area of high energy physics.

As libraries are an important part of the scholarly communication system, many library researchers have explored the OA rates of the library and information science literature. Grandbois and Beheshti [22] have studied the open availability of LIS articles and found that

60% of these are published openly. However, they criticized the relatively low rate of self-archiving, compared to the 98% option that the journals provided to the authors; in their words, this signifies “a large gap between theory and practice which needs to be addressed.” Schultz [23] has explored if researchers on OA “practice what they preach” and in her study a slightly over 25% of the articles are paywalled, which leaves an impressive 75% in various OA modes.

While the citation advantage of OA papers has been exhibited in numerous research studies [15–17]. Björk and Solomon [27] supported the claim that there are similar citation rates between OA and paywalled works. Similarly, Craig et al. [28] have articulated a critique towards this claim and support that the main effect is not attributed to access, but to discovery of sources. On another critique however, Lewis [29] supported that pirate access (Black OA) to paywalled publications is more likely to degrade this advantage and sustain the current status of citation performance. Nevertheless, recent large scale studies [30] have exhibited significant differences in citation indicators, like RCR [31], that reach up to 130% between the citations of OA and paywalled articles. Huang et al. [32] have provided evidence that the OA status of medical publications is not only responsible for increased citations, but consequently benefit the Impact Factor of the journals that they are published in. Evans and Reimer [33] have underlined the fact that the combination of traditional publishing services and open access reinforces citation advantage, compared to free online and early publication.

### 3. Research Questions & Methodology

In this study we explore the Open Access and Open Science literature, a topic-defined set of publications, as published in a period of twenty years, from 1999 to 2018, and indexed in the bibliographic database of Web of Science. While the two topics have differences in scope, we preferred to focus on both, as a representation of the two broadest topics for the current trends and changes in the scholarly communication field. Our exploration of the literature is channeled into two directions:

1. firstly, what can be revealed by the macroscopic view of the access status of major entities, namely publications, authors and sources, and is there evidence of citation advantage of openly accessible publications in this field?
2. secondly, can a microscopic view of its temporal evolution exhibit attributes of researchers’ practices and behaviors?

We searched in all Clarivate Analytics’ Web of Science indexes, namely SCI-EXPANDED, SSCI, A&HCI and ESCI for Open Science and Open Access publications. We

limited the scope of the query to the ‘Title’ field and after consecutive attempts we finalized our query string using Boolean operators to avoid homonym, but semantically different terms. The term ‘open access’ is common for other research fields too, including medical operations (open access endoscopy), rural development (open access land), fishery (open access fishery plants), and so on. The results were narrowed to those of the years 1999-2018 and the initial export was 3.042 records, that -after further cleaning- resulted to 2.846 records. These records were pre-processed to remove redundant columns and to check for syntactic consistency and error-prone characters. Frequent values in certain columns were coded to integers to enable accurate processing. during of which, certain variations were grouped to large categories; an example being the Gold Open Access option which also included the ‘DOAJ Gold’ records. Analyses were performed in Python, using Apache Zeppelin Notebooks, a web-based notebook environment that enables interactive big data analytics.

## 4. Findings

### 4.1. Statistics

#### 4.1.1. Publications

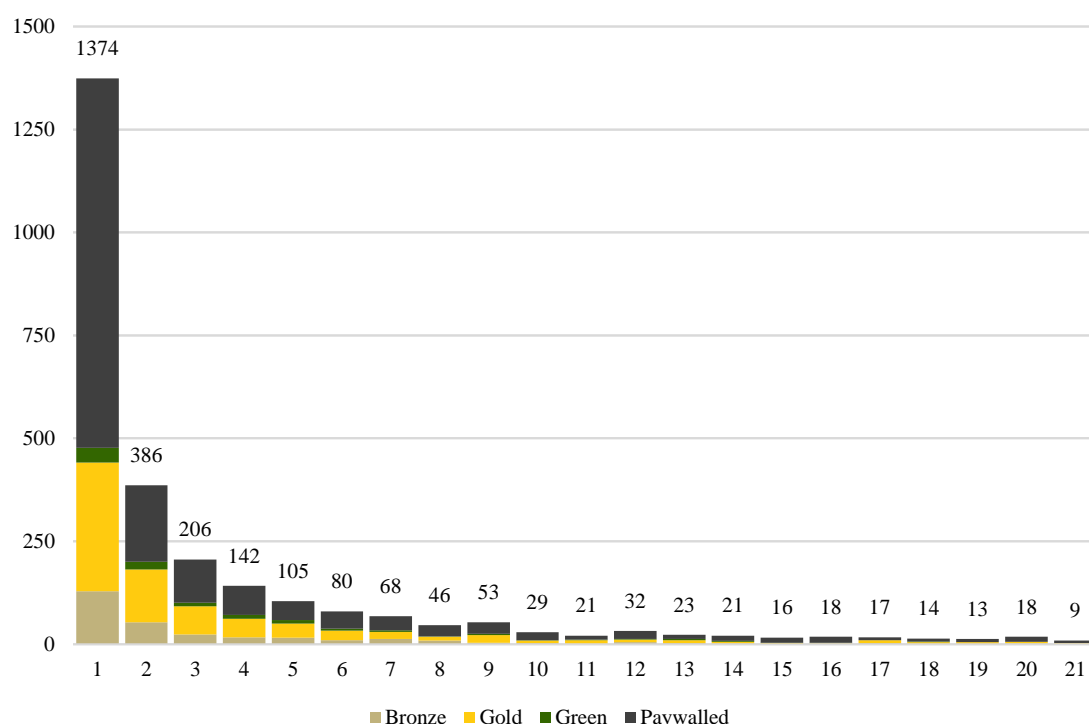
A publication equals to an entry in the TI column of our dataset. Each title is a publication, but there are cases where two or more publications share the same title. Such is the case of the title ‘Open Access’, which is common for 18 distinct publications.

These publications are grouped to seven Document Types, namely Article, Correction, Editorial, Letter, News, Review (of book, conference, etc.) and Unlabeled, with the latter grouping smaller representations of other document types. Article is the main type in our data (n=1.184, 42,5%), followed by Editorial notes (n=908, 31,9%). At the same time, there is strong representation of Letters to the Editors (n=218), while the popularity of OA/OS issues is also reflected on a significant amount of News items (n=223). Until 2004 the main contribution types appear balanced. In fact, in 2004 editorial notes are more than research articles. Research articles have a significant growth after 2015, while the consistent presence of Letters to the Editors after 2000 is noteworthy. Since 2014 there is an important growth of Editorial notes, with 2015 recording an impressive record of 100 notes.

The publications are categorized into four Access Types; three of them, Gold, Green and Bronze, gathered all aspects of OA types, while Paywalled represented the toll type of access. In our data, almost 57,5% of the publications are behind paywalls (n=1.634), while Gold

represents approximately 27% (n=772), followed by Bronze (n=312) and Green (n=128). Paywalled publications have significantly increased after 2012 (52,17% since 2011) and remain to these levels with a tendency to further increase. The respective growth for Gold OA papers is recorded in 2014. Bronze publications have been consistent in all years since 2003, while Green shows substantial numbers after 2015.

Figure 1 shows the number of citations that the publications of each year have received until the data collection time point. In our dataset there are 1.374 publications (48,28%) that have never been cited, and 386 that have been cited only once. At the same time, there are 80 papers that have been cited 5 times, 21 papers cited 10 times and nine papers that have been cited 20 times.

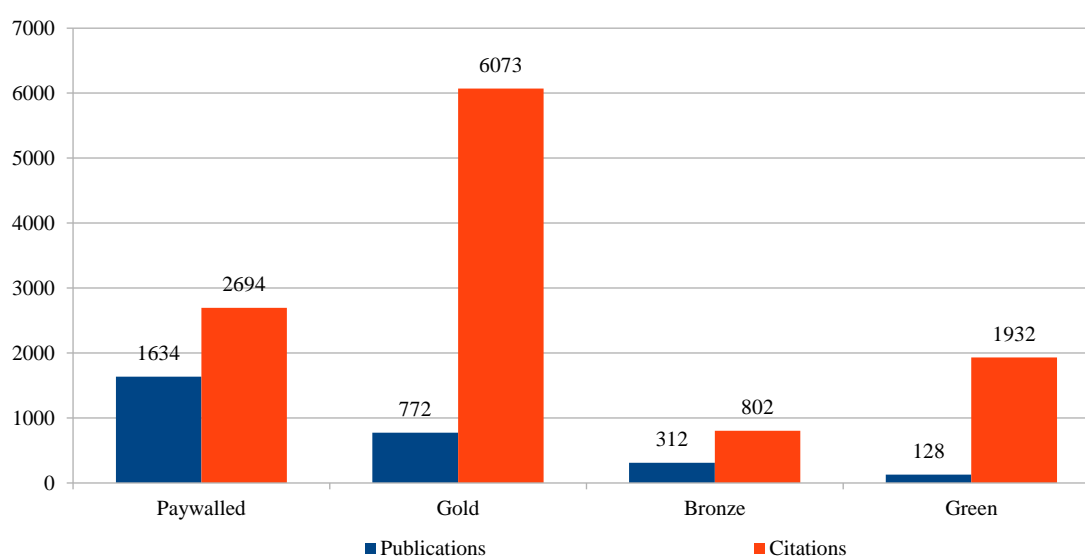


**Figure 1.** Number of times cited per year

The most cited paper, an annual report of an open access database, has received 855 citations and it is responsible for the high number of citations in 2004, the year of its publication. Similar is the case of the year 2008, when there is a new version of this report that accounts for 467 citations, while both years have the highest average of times cited, see 14,18 and 14,01 respectively. There are also three years - 2006, 2009 and 2010 - when the average citation number exceeds ten citations per publication, see 10,20, 10,15 and 10,70 respectively.

## The case of citation advantage

The ‘citation advantage’ is a claim that has been supported by many OA advocates, which refers to an increased number of citations to OA publications, and, therefore, we examined whether there can be differences in the citation patterns of each Access Type in this set of OA/OS publications. Figure 2 shows that there is evidence of citation advantage for all OA publications and most notably for those published in the Gold Road. Gold publications which represent 27,13% of the overall number have received 6.073 citations that corresponds to 52,8% of the overall citations. Similarly, Bronze and Green represent 6,97 and 16,79 of the overall citations, while the citations produced by the paywalled publications, which correspond to 57,41%, is 23,42%. Thus, the average number of citations per each Access Type is 1,65 for paywalled, 2,57 for Bronze, 7,87 for Gold and 15,09 for Green publications. It must be noted that Green publications are enforced by that one publication that has received 855 citations alone.



**Figure 2.** Number of publications and number of times cited per access type

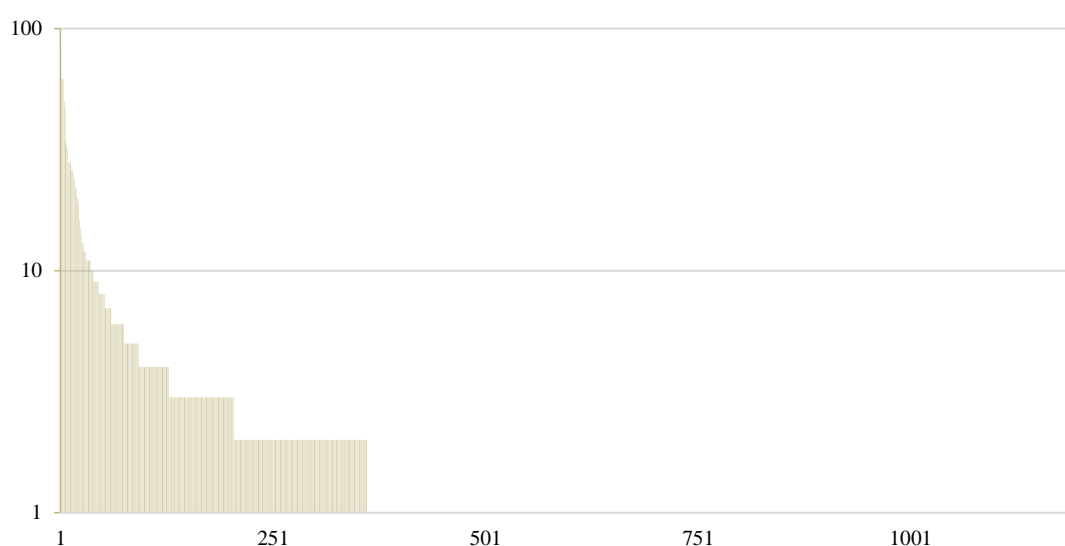
### 4.1.2. Sources

In the WoS nomenclature Sources is a term that describes publication venues, which, most of the times, corresponds to scientific journals. There were 1.197 distinct sources in our dataset. Histogram inspection showed us that most of these sources ( $n=1.153$ ) have published up to three articles on this topic, while on the other end only two sources have published from 9 to 11 articles. Figure 3 presents the histogram of the number publications per source. The bins on the x-axis represent intervals of numbers of sources and the corresponding value on the y-axis (in logarithmic scale) is the number of publications which have been published in



the sources of each bin. In general, 87 sources have allowed Green deposition of their publications, 162 journals are associated with the Bronze publishing mode, and 357 have published articles in the Gold OA road. The majority of the journals though (n=711) have published articles in toll access mode.

From a list of the 35 most frequent sources, that had published more than ten articles each on this topic, we manually checked through the JournalTOCs service,<sup>1</sup> which of those are Open Access, Hybrid or Paywalled sources. Fifteen sources, that publish 48,5% of the publications of these, are hybrid, meaning that they provide the OA option to authors. Eleven of the sources are entirely Open Access, with two of them being ‘free’, while nine of them are behind tolls. The OA sources have published 203 works (24,1%) and the paywalled 261 (27,4%), while the greater part of this list belongs to the Information & Library Science discipline, a category followed by Multidisciplinary journals (n=5) and Medical Science journals (n=4). Journals from Chemistry and Physics follow in the ranking, while the other two belong to Biology and Computer Engineering.



**Figure 3.** Histogram of publications per sources

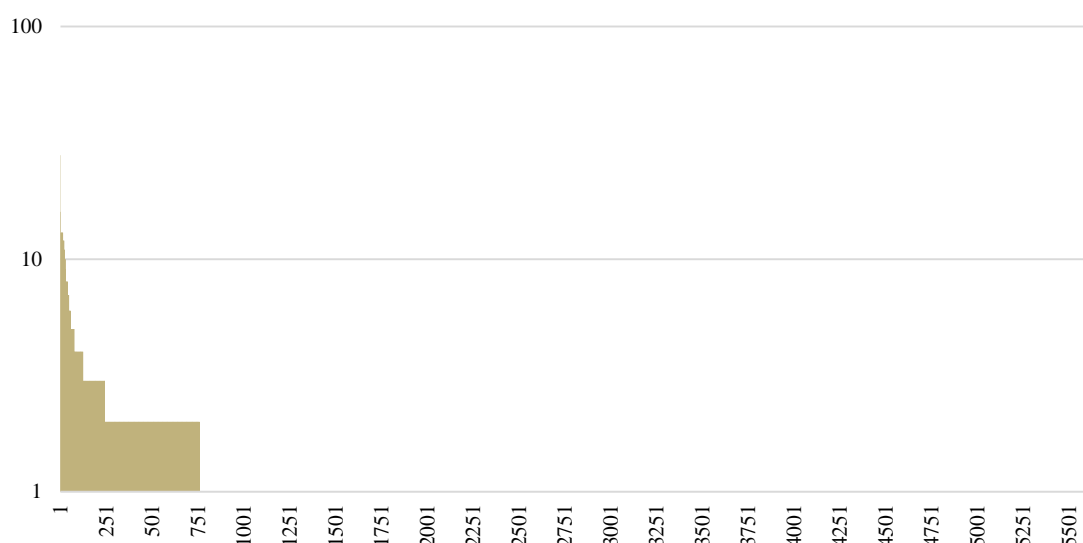
Since 2012 the number of sources that publish items on this topic is consistently increasing. The last five years, from 2014 to 2018, the increase of sources corresponds to 75,35% (from 142 to 249 sources). The first five years of the period under question, the growth was rather weak and the interim stage from 2004 to 2011 was moderate, which means that 50 to 100 titles per year were publishing relevant papers. These sources are engaged into

<sup>1</sup> <http://www.journaltoCs.ac.uk>

the discussion in various ways and not just in publishing research articles. Interestingly, more source titles have published more Editorial Notes than research articles, which is the prime type of contributions of the dataset. Reviews have been published in 93 sources, while the Editors of 80 journals have received and published Letters from readers. The significance of the field is also reflected to the 53 journals that have reported news items.

#### 4.1.3. Authors

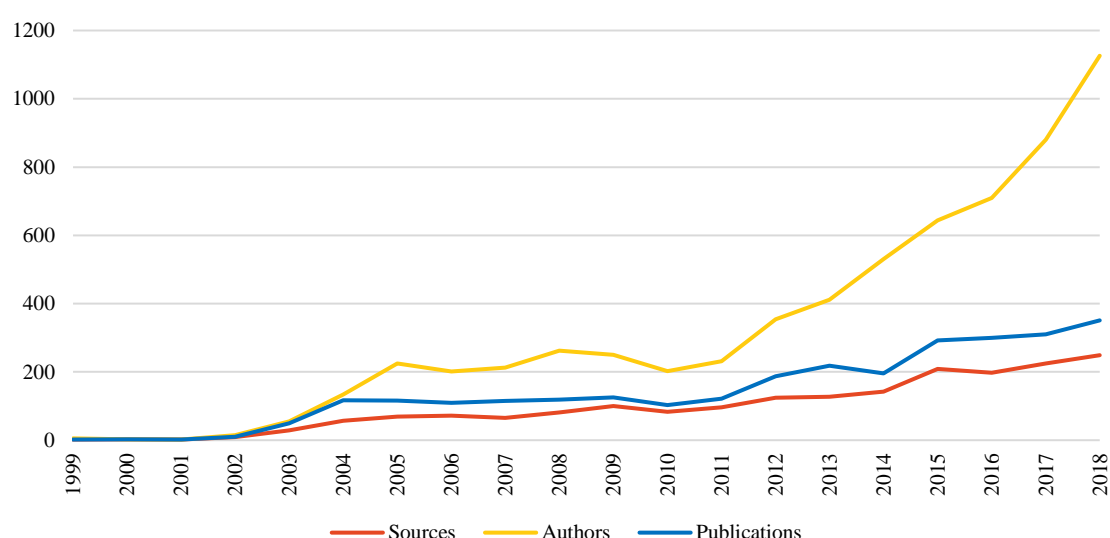
An authoring entity is every value in the AU column of our data. Therefore, the 2.864 publications of our data have been authored by 2.362 authoring entities, that range from 1 to 138 co-authors in a work. When split and deduplicated, these entities result to 5.659 individual authors, which equals to 11,22 authors per publication. Figure 4 presents the histogram of publications per author. The average number of authors in each authoring entity is reaching the number of three in 2014 and 2017, see 2,99, and overpasses it in 2018 with 3,59 authors per publication.



**Figure 4.** Histogram of publications per author (y-axis in logarithmic scale)

From 2005 to 2011 the number of individual authors remains stable, approximately 250 authors per year. This number is increased after 2012, surpassing the limit of 1.100 authors in 2018. This is obviously related to the growth of publications and the number of authoring teams that contribute, both of which are the highest during that year. Indeed, the average number of authors per publication in this year is 3,21, while during the last five is steadily above the average number of two authors per publication.

It is impressive that in 2014 there are more authors publishing in Gold OA items, than in Paywalled. Despite the fact that the highest average number of authors per Gold publications is in 2008 (4,79), in 2014 we record the second highest number, see 4,59. For an entire decade, from 2007 to 2016, there are more authors in Bronze than in Green publications. However, the last three years, the number of authors who publish their works in Green OA is increasing, see 5,17 in 2016, 6,19 in 2017 and 12,35 in 2018, showing that a culture of multi-authorship is growing among authors who prefer the Green road. Figure 5 presents the development of the growth indicators for Authors, Sources and Publications in this period.

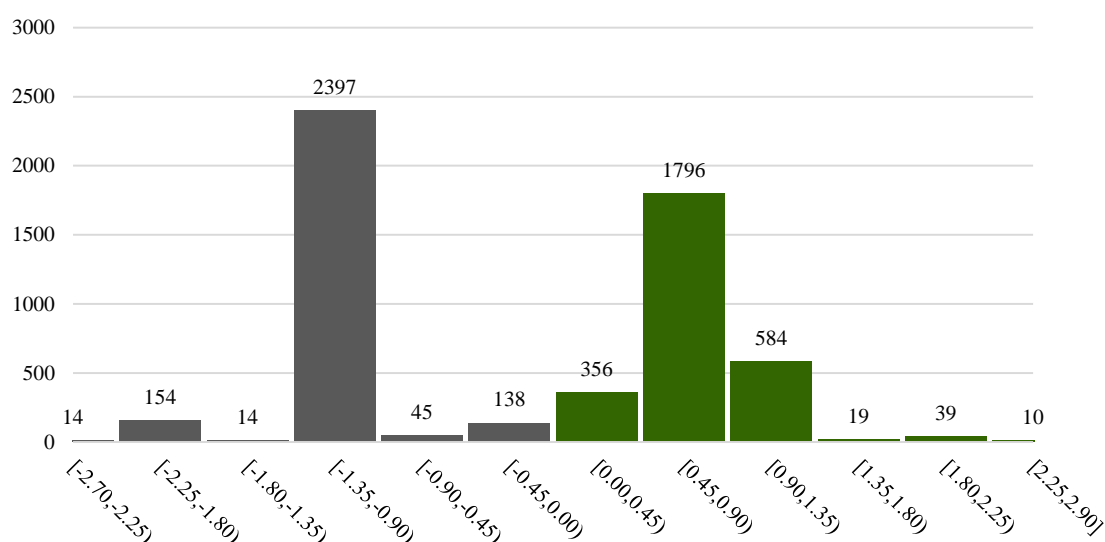


**Figure 5.** Growth of Sources, Authors and Publications

### *Openness of Individual Authors' Publications*

Many have questioned how 'openness of a researcher' can be estimated [34, 35]. Considering just a single aspect of openness, i.e., the authors' publication production, we have managed to calculate a simple indicator of openness. We have assigned a weighted score to the Access Type of every publication of every author and then we have counted the scores of every publication. The weight scores that we have used were -1 for Paywalled, 0,3 for Bronze, 0,7 for Gold and 1 for Green publications. This scale was devised based on the principle of the economic aspect of OA, meaning that we have opted for Green at the most positive end, as a cost/benefit model that secures OA through existing infrastructures controlled by academic and research communities [36]. Figure 6 displays the histogram of

the number of authors in subintervals of ‘openness’ between -2,7 and 2,9 (excluding three outliers). Notice that there is a balance between authors who have a positive (n=2.804) and a negative (n= 2.762) score. Given the fact that most of the authors have published one to two papers, the rates are close to 0. This would also explain the large concentration of authors’ openness scores in the three bins of the positive and in the third bin of the negative end.



**Figure 6.** Highest and lowest scores of author openness based on the access type of publications

#### 4.1.4. Other dimensions of OA/OS publications

In our data, publications have been classified into seven languages. Most of them is authored in English (n=2.583), confirming that this is the main language of scholarly communication. In fact, during the first five years, the literature, although limited, is published entirely in English, with a boost of the publications in this language in 2004 and onward. English are followed by Spanish (n=92) and German (n=87), while the rest of the order, as per their magnitude, is Portuguese, Italian, French and Russian. A final option, ‘Other’, was used to group other -fewer- instances of languages. The only year that there were publications in every language is 2017.

Whenever we found a funding note in a record, we coded it and sorted it in two categories, ‘With Funding’ and ‘Without Funding’. Only a small number of these publications have this kind of note (n=351), which leaves the majority as not funded (n=2.945). In the first five years of these two decades, the occasion of a funded article is very rare. The first funding note has been traced in 2004, while increased numbers are recorded from 2014 and onward.

Descriptors are keywords that the authors themselves have chosen to describe their publications. There are 808 distinct descriptors, but there are no descriptors at all for 2.031 papers. At the same time, WoS assigns its own keywords to publications, Identifiers, or else Keywords Plus®. Again, a significant part of the publications (n=2.045) are not assigned with an Identifier. Table 1 provides the list of the ten most frequent Descriptors and Identifiers, excluding ‘None’, with their respective percentages.

**Table 1.** Top ten Descriptors and Identifiers<sup>1</sup>

<b>Descriptors</b>	<b>%</b>	<b>Identifiers</b>	<b>%</b>
Open Access	23,95	Impact	7,34
Open Science	4,33	Science	5,48
Scholarly Communication	2,63	Journals	4,76
Publishing	2,63	Articles	1,69
Open Access Journals	2,01	Communication	1,69
Institutional Repositories	1,82	Web	1,63
Open Access Publishing	1,69	Information	1,63
Journals	1,57	Authors	1,63
Scholarly Publishing	1,44	Model	1,56
Repositories	1,25	Publication	1,50

<sup>1</sup> Since these are author generated keywords, some are variations of the same concept.

## 4.2. A temporal micro analysis of practices

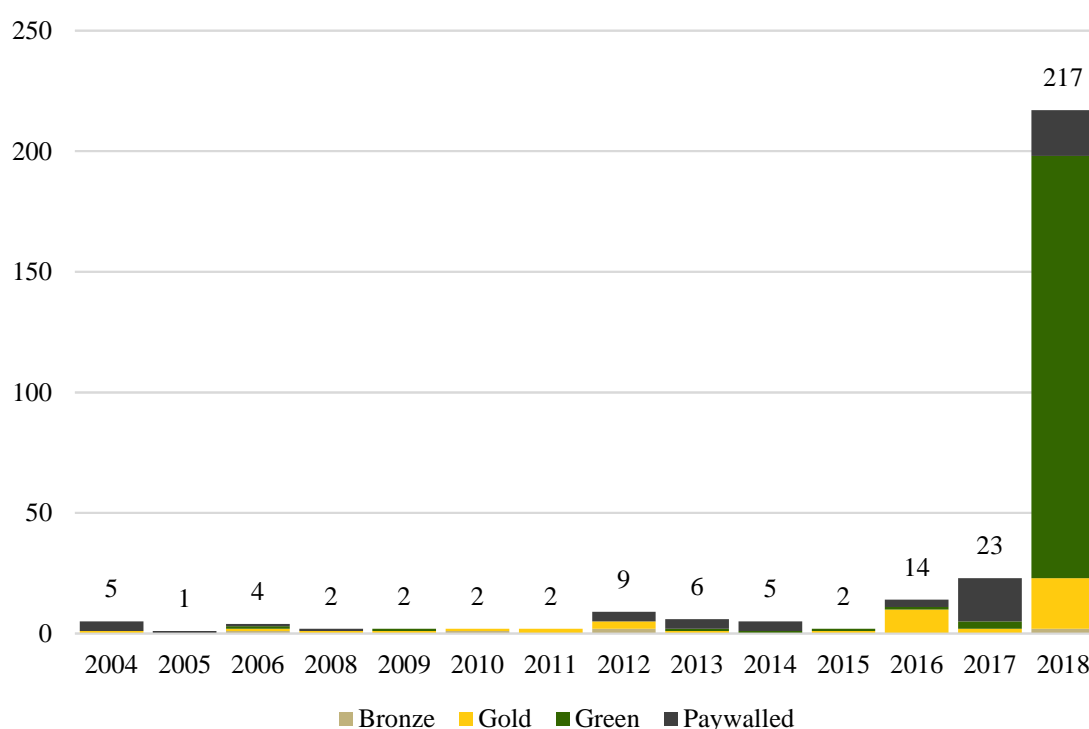
In this section we explore the practices and the behaviors in the temporal evolution of the field. In particular, we split our analyses to (i) Practice Variations, which analyzes the publication practices of particular authors and sources that have been qualified for this based on their activity, (ii) Population Behavior, which concentrates on the consistency of Authors, both in general and per Access Type, and the expansion rate of the field, and (iii) Publication Loyalty, which explores microscopically the levels of consistency can be traced in publication patterns.

### 4.2.1. Practice variations

### *Variations for particular authors*

Out of the 5.659 individual authors we selected the following 22 based on the threshold of more than ten publications by each, excluding ‘Anonymous’. The earliest record by this group of authors can be found in 2004, yet the main activity is recorded in the last three years. There are 285 publications that, in the wide selection of journals, have been published in 36 unique sources. The number of sources, following the trend line of the publications, is consistently low for most of the years, having a rapid increase in the last three years. Checking in the WoS categories the disciplines of these sources revealed that most of them belong in the Medical Sciences (15 sources) and Library & Information Science (six sources), while 31 of these have been ranked in the Journal Citation Index.

As outlined earlier, the consistency of publication is rather disturbed the last three years, and this is evident also in the Access Types of the publications of these authors. The boost in 2018 has resulted to a large number of authors in Green publications and despite the fact that the authors with Bronze publications are few in numbers, the other two categories of OA have higher scores, even than Paywalled publications. This small anomaly is attributed to 13 publications by an authoring entity of 23 authors, with varied composition, that has been published with the same title in 13 different journals in 2018. In the majority of these years, the top 22 authors that write and comment on OS and OA publish paywalled publications in almost every year.

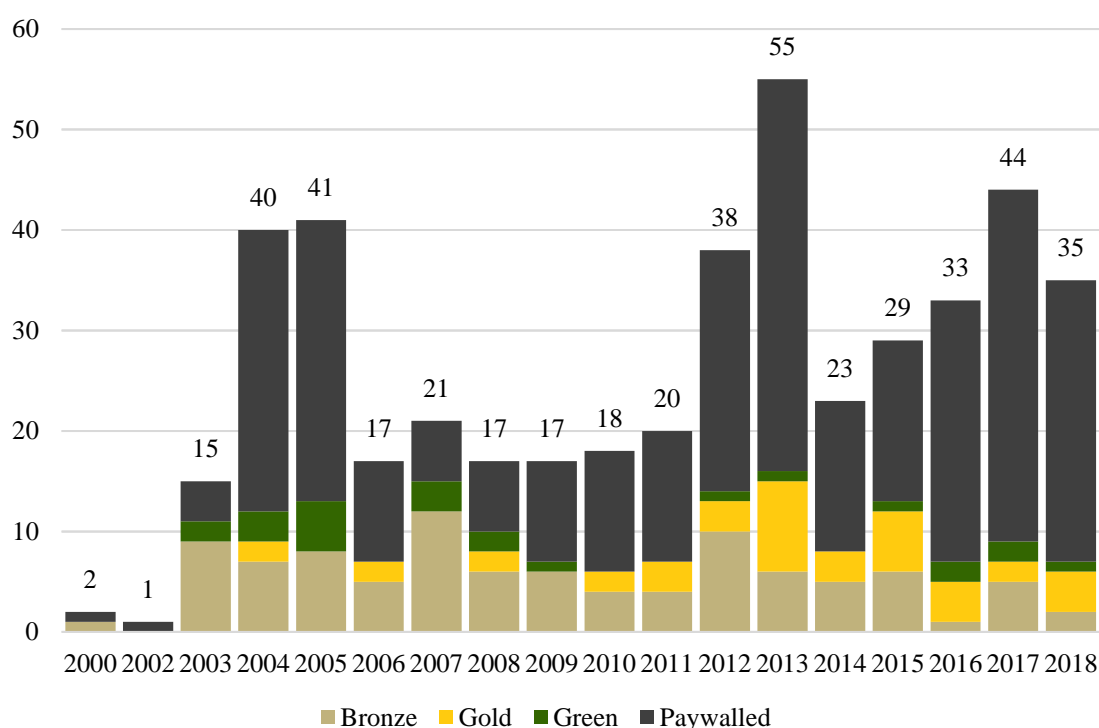


**Figure 7.** Variations of top authors' publications per access type

We also examined the variations of Document Types of these authors, which during the entire period reflect a diverse landscape of contributions. The only year that there is invariability is the last one, 2018, when we have a boost in editorial notes, followed by research articles. There has also to be noted that in 2017 there is a relatively high number of Letters to the editor by these authors, that even surpasses the number of articles.

#### *Variations in particular sources*

We reversed our point of view and selected the top nine sources that have hosted more than 30 publications each. The nine sources are: British Medical Journal, Learned Publishing, Chemical & Engineering News, Nature, Abstracts of Papers of the American Chemical Society, Current Science, Journal Of Academic Librarianship, Scientometrics, and Science. Seven out of these sources are also in the list of 36 journals that the 22 most active authors have published in. These journals are supported by the publication of 515 unique authors, with the years of 2013 and 2017 having attracted more, 93 and 66 respectively. Out of these 515 authors, there are five who belong to the group of top authors in the previous section.



**Figure 8.** Variations of top sources' publications per access type

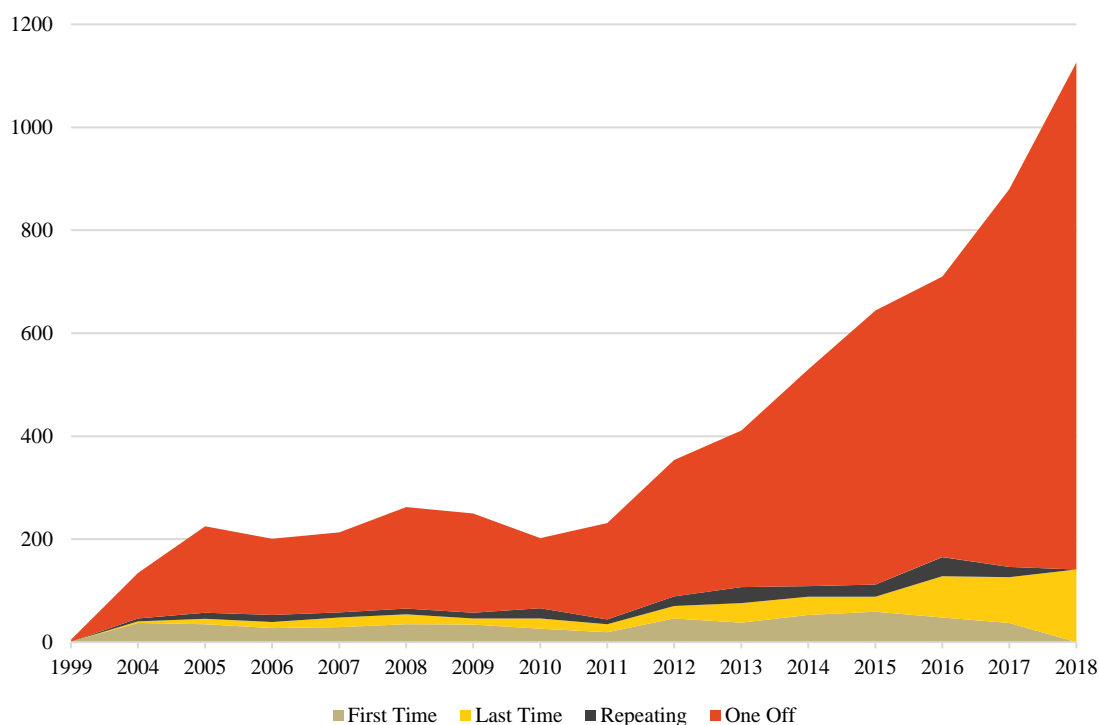
Figure 8 shows what is the access type of the publications in these journals during the years. As most of these nine journals have subscribed or hybrid identity, it is reasonable that we meet paywalled publications in every year. There percentages are always the highest ones, except for 2007 when Bronze publications have taken the lead in this ranking. Since 2010, Gold OA publishing are consistently present in these journals.

Research articles and news items are the most popular types of contributions. Articles have been published in considerable amounts in the Journal of Academic Librarianship in 2013, in Learned Publishing in 2017 and 2018, and in Scientometrics in 2018, while news items have been published extensively in BMJ and Chemical and Engineering News in 2013 and in Science in 2018. Editorial notes have been published in Nature in the last two years. What it can be alarming for the publishing and indexing services is that the scores of Unlabeled contributions are quite high and almost omnipresent, as it happens in the case of Abstracts of Papers of the American Chemical Society.

#### 4.2.2. Population behavior

Every research field is developed over time and authors are the driving force as through their publications help to shape its understanding and direction. Therefore, we examined the presence of authors through time, in order to understand if the scholars of this field are consistently interested in it. To do so, we group authors in four categories: (a) First Time Authors are those who first appear in a publication in a year and while appear again in any later year of our dataset, (b) Repeating Authors are the ones that appear in publications in consequent years, (c) One Off Authors appear in a publication once, but they do not appear again, and (d) Last Time Authors are the ones that have appeared in any past year, but they appear for the last time in a specific year of our dataset.





**Figure 9. Population behaviour<sup>2</sup>**

<sup>2</sup> In 1999 there can be no Last Time and Repeating Authors, while in 2018 can be no First Time and Repeating Authors.

Figure 9 shows the general time development of the field, which is represented generally by One Off authors. The percentages of their presence are always higher than 65,7%. On the other side, the category that shows a persistence in research publications, the Repeating Authors, have quite low rates, which has a minimum of 2,3% (2017) and a maximum of 9,9% in 2010. Many of the Repeating Authors have published a considerable amount of Bronze contributions in 2013 (10,5%), while a similar consistency by Repeating Authors is recorded for Gold contributions in 2006 (11,6%). The rate of Repeating Authors in Paywalled articles is quite low (from 2,5% to 5,8%), while the number of First Time is decreasing the last five years.

Having this information available we calculate an expansion rate, which is the count of the One Off and the First Time Authors to all authors in a year. This allowed us to estimate the streams of new authors during time. 2009 and 2015 are the years that new authors entered the publication field, with a rate of 0,91 and 0,92 respectively. In 2019, the majority of new authors seem to prefer publishing paywalled items, while in 2015 the new authors are responsible for the main contributions.

#### 4.2.3. Publication Loyalty

In sub-section 4.2.2 we approached the concept of consistent publication behavior. Zooming in the publication history of these sources we further explored the consistent presence, or else continuity pattern, of authors. Two authors have published for four consequent years in the same journal, followed by a group of other authors with three publications in subsequent years. We also explored the notable periods of absence from the same source, meaning how many years of lag have been recorded between two publishing instances. In this case, two authors had a considerable amount of years before returning to publish in a source they had published before, one with nine years and another with eight years, followed by a group of authors with seven years of absence.

We produced a similar analysis for the sources that had published works of the same authors for many years. There ten journals that have hosted three publications in a row by authors, while *Physics World* and *Chemical Engineering Letters* have published for four consequent years their contributions. The lapse pattern for sources has similar periods, as the majority of the journals (six) have seen authors returning to publish after seven years, while there are journal titles, like *Nature*, *PLOS Medicine* and *Learned Publishing* that have eight or more years to host the publications of previously appeared authors.

## 5. Discussion

The field of OS/OA is rapidly evolving the last five years, although substantial growth has been recorded since the start of the second decade. This is in accordance with other studies that focus on the growth of OA documents in general, but also holds similarity to other research topics, like Research Data Management [37]. This shows that the tectonic plates of scholarly communication are on the move, shaping a different landscape though intense communication. The reasons of course can be multifaceted and it can be attributed to the increased interest on the field, as well as to commentaries on developments, such as the NHS policy in 2008, the various subscription cancellations in US and EU Universities in 2017 and lately the PlanS implementation plan. Even though there are differences in the OA representation, our findings on the growth in OA/OS publications are similar to Schultz's, that have indicated 2013 as a strong turning point for their increase. In our study, the majority of publications is still in paywalled mode, but the OA options gradually become more competitive. In the case in point, the Gold OA road represents an option that more than one quarter of the publications have preferred, especially in the last five years. This is however in contrast to other studies like Gargouri et al. [20] who have evidence of increased availability of Green OA publications over Gold ones.

There is evidence that the field is expanding not just in absolute publication numbers, but also in terms of authors that enter and join the existing ones. However, combining evidence from the long tail of author contributions and the population behavior, we can understand that the field is having a greater part of authors with limited connection. The rate of authors who publish consistently in the field is quite low, ranging from 5 to 10% and these ‘returning’ authors prefer the Gold road to express, especially during the second decade of the period. On the other side, authors who publish paywalled items seem to do so scarcely and inconsistently. The Gold road also exhibits a consistent rate of new authors every year, in contrast to the other access types, including toll access. The bonds between authors and sources seem to be loose, as there are only a few authors that publish in the same source for several consecutive years.

More journals are taking part in the dialogue. While the library and information science journals have a strong interest in this topic, the discourse is not their own benefit. Multidisciplinary journals, as well as journals from chemistry and physics are platforms for communicating opinions and findings. A sample of the journals in our dataset revealed that the majority are hybrid venues, while only a small amount is entirely OA. The road to full OA is still long, and as recent author surveys have shown the prestige of the journal or its esteem in the community [27, 28] are more important reasons to select a publication venue than the OA status.

Based on the current data, we proposed an exercise for an openness metric that relates to the publication activity of a researcher. This metric is not of course comprehensive, as the concept of openness envelops other aspects as well, as it has been proposed in OS-CAM [40] and Open Access Spectrum Evaluation Tool [41]. Yet, based on the intrinsic properties of publications one can have an estimation of how open the work of an author is. Adjustments on the inclusion of alternative versions of the publication of an author can consider the multi-variety of publishing options.

With a sizable dataset and microscopic exploration of the behaviors, we were able to reveal certain practices that disturb the normality. Versioned documents that address to concrete and active communities can result in high citation scores, while bursts of publications by large authoring teams can magnify author productivity. Despite these small anomalies, the citation advantage claim has a solid grounding in our study, as there are ample differences between the Gold OA and the paywalled publications. Although one can exclude from the discussion the citations of the last two years, there is a recession in the number of

citations per year, which can be attributed to the increase of peripheral document types, such as Editorial notes and Letters to the Editors, that regularly do not receive many citations.

Inevitably, our study has certain limitations including the use of a bibliographic database to retrieve our data. Much of the discourse is published on other means than scientific journals, such as trade journals, reports, working documents and statements, while not all journals from all countries are listed in this database. Therefore, the literature that is indexed has been subjected to selective policies that cannot reflect inclusivity.

## 6. Conclusions

Open Science and Open Access are two topics that have concerned heavily the researchers, as they focus on fundamental changes in scholarly communication that our world has not witnessed for many years. For these changes to be as effective as possible, informed decisions should be based on every available information and knowledge resource. Currently, the state of the OA/OS literature is showing signs of growth, but it still needs to be released from tolls and reach wider audiences. While the advantages of OA are known to everyone, and the citation advantage has been confirmed also in this study, this literature is still being published in subscription venues.

## References

1. Tennant JP, Waldner F, Jacques DC, et al (2016) The academic, economic and societal impacts of Open Access: an evidence-based review [version 3; peer review: 4 approved, 1 approved with reservations]. F1000Research 5:. <https://doi.org/10.12688/f1000research.8460.3>
2. LIBER (2018) LIBER Open Science Roadmap
3. LERU (2018) Open Science and its role in universities: a roadmap for cultural change
4. Priem J [@jasonpriem] (2018) Wow. Did not realize that this entire special issue on #openaccess is...paywalled. Can read whole thing for low price of \$300. Words fail me
5. Gutknecht C [@chgutknecht] (2019) A report about the progress of #openaccess agreements in Switzerland, which is closed access? Seriously?? <https://doi.org/10.1515/abitech-2019-3002>

- 535 6. McKiernan EC, Bourne PE, Brown CT, et al (2016) How open science helps researchers  
536 succeed. *Elife* 5:e16800. <https://doi.org/10.7554/eLife.16800>
- 537 7. European Commission (2019) Open science monitor. [https://ec.europa.eu/info/research-](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)  
538 [and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)  
539 [science-monitor\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)
- 540 8. SPARC Europe, Digital Curation Centre (2019) An Analysis of Open Science Policies  
541 in Europe v4. In: Zenodo
- 542 9. Rodrigues SR, Taga V, Passos FM (2016) Research Articles about Open Access Indexed  
543 by Scopus: A Content Analysis. *Publ.* 4
- 544 10. Björk B-C, Welling P, Laakso M, et al (2010) Open Access to the Scientific Journal  
545 Literature: Situation 2009. *PLoS One* 5:e11273
- 546 11. Laakso M, Welling P, Bukvova H, et al (2011) The Development of Open Access  
547 Journal Publishing from 1993 to 2009. *PLoS One* 6:e20961
- 548 12. Archambault É, Amyot D, Deschamps P, et al (2014) Proportion of open access papers  
549 published in peer-reviewed journals at the European and world levels—1996–2013
- 550 13. Piwowar H, Priem J, Orr R (2019) The Future of OA: A large-scale analysis projecting  
551 Open Access publication and readership. *bioRxiv* 795310.  
552 <https://doi.org/10.1101/795310>
- 553 14. Piwowar H, Priem J, Larivière V, et al (2018) The state of OA: a large-scale analysis of  
554 the prevalence and impact of Open Access articles. *PeerJ* 6:e4375.  
555 <https://doi.org/10.7717/peerj.4375>
- 556 15. Fathli M, Lundén T, Sjögarde P (2014) The Share of Open Access in Sweden 2011 -  
557 Analyzing the OA outcome from Swedish Universities. *ScieCom Info* 10:
- 558 16. Rovira A, Urbano C, Abadal E (2019) Open access availability of Catalonia research  
559 output: Case analysis of the CERCA institution, 2011-2015. *PLoS One* 14:e0216597
- 560 17. Breugelmans JG, Roberge G, Tippet C, et al (2018) Scientific impact increases when  
561 researchers publish in open access and international collaboration: A bibliometric  
562 analysis on poverty-related disease papers. *PLoS One* 13:e0203156
- 563 18. Kurata K, Morioka T, Yokoi K, Matsubayashi M (2013) Remarkable Growth of Open

- 564 Access in the Biomedical Field: Analysis of PubMed Articles from 2006 to 2010. PLoS  
565 One 8:e60925
- 566 19. Anan H, Liu X, Maly K, et al (2002) Preservation and transition of NCSTRL using an  
567 OAI-based architecture. In: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on  
568 Digital Libraries. ACM Press, New York, N.Y., p 181
- 569 20. Gargouri Y, Larivière V, Gingras Y, et al (2012) Green and Gold Open Access  
570 Percentages and Growth, by Discipline. In: Archambault É, Gingras Y, Larivière V (eds)  
571 Proceedings of 17th International Conference on Science and Technology Indicators,  
572 Montréal: Science-Metrix and OST
- 573 21. Morison H (2018) 2018: best year yet for net growth of open access.  
574 [https://poeticeconomics.blogspot.com/2018/12/2018-best-year-yet-for-net-growth-](https://poeticeconomics.blogspot.com/2018/12/2018-best-year-yet-for-net-growth-of.html)  
575 [of.html](https://poeticeconomics.blogspot.com/2018/12/2018-best-year-yet-for-net-growth-of.html)
- 576 22. Grandbois J, Beheshti J (2014) A bibliometric study of scholarly articles published by  
577 library and information science authors about open access. Inf Res 19:
- 578 23. Schultz T (2018) Practicing What You Preach: Evaluating Access of Open Access  
579 Research
- 580 24. Hajjem C, Harnad S, Gingras Y (2006) Ten-Year Cross-Disciplinary Comparison of the  
581 Growth of Open Access and How it Increases Research Citation Impact
- 582 25. Eysenbach G (2006) Citation Advantage of Open Access Articles. PLOS Biol 4:e157
- 583 26. Swan A (2010) The Open Access citation advantage: Studies and results to date
- 584 27. Björk B-C, Solomon D (2012) Open access versus subscription journals: a comparison  
585 of scientific impact. BMC Med 10:73. <https://doi.org/10.1186/1741-7015-10-73>
- 586 28. Craig ID, Plume AM, McVeigh ME, et al (2007) Do open access articles have greater  
587 citation impact?: A critical review of the literature. J Informetr 1:239–248.  
588 <https://doi.org/https://doi.org/10.1016/j.joi.2007.04.001>
- 589 29. Lewis DW (2016) Will Sci-Hub Kill the Open Access Citation Advantage and (at least  
590 for now) Save Toll Access Journals? In: IUPUI Sch. Repos.  
591 <https://scholarworks.iupui.edu/handle/1805/11161>
- 592 30. Wizdom (2018) Open Access - Global Trends and Rankings.

- 593 <https://www.wizdom.ai/dashboards/open-access>
- 594 31. Hutchins BI, Yuan X, Anderson JM, Santangelo GM (2016) Relative Citation Ratio  
595 (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article  
596 Level. *PLOS Biol* 14:e1002541
- 597 32. Huang C, Yue X, Chen J, et al (2019) The effect of “open access” on journal impact  
598 factors: A causal analysis of medical journals. *Phys A Stat Mech its Appl* 533:122043.  
599 <https://doi.org/https://doi.org/10.1016/j.physa.2019.122043>
- 600 33. Evans JA, Reimer J (2009) Open Access and Global Participation in Science. *Science*  
601 (80- ) 323:1025 LP – 1025. <https://doi.org/10.1126/science.1154562>
- 602 34. Levin N, Leonelli S, Weckowska D, et al (2016) How Do Scientists Define Openness?  
603 Exploring the Relationship Between Open Science Policies and Research Practice. *Bull*  
604 *Sci Technol Soc* 36:128–141. <https://doi.org/10.1177/0270467616668760>
- 605 35. Knowledge Exchange (2019) Openness Profile. [http://www.knowledge-](http://www.knowledge-exchange.info/event/openness-profile)  
606 [exchange.info/event/openness-profile](http://www.knowledge-exchange.info/event/openness-profile)
- 607 36. Houghton J, Swan A (2013) Planting the Green Seeds for a Golden Harvest: Comments  
608 and Clarifications on “Going for Gold.” *DLib Mag* 19:.  
609 <https://doi.org/10.1045/january2013-houghton>
- 610 37. Zhang L, Eichmann-Kalwara N (2019) Mapping the Scholarly Literature Found in  
611 Scopus on “Research Data Management”: A Bibliometric and Data Visualization  
612 Approach. *J Librariansh Sch Commun* 7:.. [https://doi.org/http://doi.org/10.7710/2162-](https://doi.org/http://doi.org/10.7710/2162-3309.2266)  
613 [3309.2266](https://doi.org/http://doi.org/10.7710/2162-3309.2266)
- 614 38. Nature Publishing Group, Palgrave Macmillan (2015) Author insights 2015. *Nat. Publ.*  
615 *Gr. Palgrave Macmillan* 12
- 616 39. Taylor and Francis (2019) Taylor & Francis researcher survey 2019: Researchers’ views  
617 on the present and future of scholarly communication
- 618 40. European Commission (2017) Evaluation of Research Careers fully acknowledging  
619 Open Science Practices: Rewards, incentives and/or recognition for researchers  
620 practicing Open Science. <https://doi.org/10.2777/75255>
- 621 41. Chen X, Olijhoek T (2016) Measuring the Degrees of Openness of Scholarly Journals

with the Open Access Spectrum (OAS) Evaluation Tool. Ser Rev 42:108–115.  
<https://doi.org/10.1080/00987913.2016.1182672>



© 2020 by the authors. Submitted for possible open access publication  
under the terms and conditions of the Creative Commons Attribution  
(CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).