

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

Giannis Tsakonas¹, Sergios Lenis² and Moses Boudourides^{3,4}

¹ Library & Information Center, University of Patras, Greece

² University of Patras, Greece

³ Science Division, New York University Abu Dhabi, U.A.E.

⁴ 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 in the period 1999 - 2018 and indexed in the bibliographic database Web of Science. The total amount of publications that we explore is 2,823 and our analyses run on two levels: first, in a macroscopic point of view, we explore statistic properties of the major entities, namely, publications, authors and sources, focus on their access status and investigate how these have been developed over these two decades. 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 in this period. Our study shows that the field is expanding, not just in absolute publication numbers, but also in terms of authors expressing their interest. However, one of the main findings of this study is that authors that publish on these topics keep limited connection with this field.

Keywords: Open Access, Open Science, bibliometric study, citation advantage, openness profile, publication practices

1. Introduction

Open Access (OA) has started 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. The research on OA has grown up to a point that it can be considered an autonomous field, which studies matters on transitioning from closed to open models of access to information, upgrading and securing the quality of publication venues, supporting their accordance to research assessment, and so on. As a field of research, it has its own self-defined principles for transparency, sustainability, reusability, monitoring, and so on, that might update the way that contemporary research can be performed. In a paper that summarizes the impact of OA, Tennant et al. (2016) underline the inequality and fragility of the current publishing 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 from the state of production and processing. As a natural consequence, the scholarly communication world has realized that, in order to make changes in OA, radical changes in 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 (OS). Initiatives such as the Jussieu Call on Bibliodiversity (<https://jussieucall.org/jussieu-call/>) and the San Francisco Declaration of Research Assessment (<https://sfdora.org>) acknowledge that changes in the wider ecosystem are tightly connected to the advancement of Open Science. Lately, various political initiatives and decisions, either from research funding organizations (including the European Union, Wellcome Trust, Bill & Melinda Gates Foundation, Arcadia Foundation, and PlanS) or from governments (including national acts in the Netherlands, France, Italy and others), have intensified the discourse about the quantity of the OA that should increase and the quality that should be achieved. The various roads of OA, together with the

51 policy and the infrastructure prerequisites they carry together, are shaping the future of schol-
52 arly communication. Therefore, the need for opinion exchange and informed audiences is im-
53 perative and this need is expressed as a growing part of the literature that discusses aspects of
54 OA and OS.

55 As long as cultural traits of the existing paradigm are in force, OS is impeded and many of
56 its advocates are calling for a change of culture that will truly enact principles of increased
57 transparency, openness, integrity, accountability and so on. In order to exemplify the benefits
58 of OS, scholarly communication actors will need to commit to its principles and adopt its prac-
59 tices consistently and robustly (LERU, 2018; LIBER, 2018). However, several contradictions
60 are being recorded. Mentions about articles that discuss OA developments but are published in
61 toll-access platforms of emphatically criticized commercial publishers are feeding the social
62 media timelines, resulting in comments of wry nature (Gutknecht, 2019; Priem, 2018).

63 The topics of this debate are highly complex and require thorough examination. However,
64 after twenty years of research, there is a critical mass of literature and its characteristics, such
65 as the degrees of openness or its scholarly formulations, have not been explored thoroughly. In
66 this study, by examining bibliographic records, we are looking to understand how this literature
67 has progressed over a period of twenty years We perform a quantitative analysis of 2,823 pub-
68 lications that focus on OS and OA, as indexed in the Web of Science, to explore the develop-
69 ment of this literature over the last twenty years and we question how much of this literature is
70 open and how major entities have been developed over these years. We are interested in the
71 ways the interest of the community has been expressed through features that form the scholarly
72 outputs, such as the variations per access or per document type, as well as exploring whether
73 other questions of the topic, such as the citation advantage claim, can be answered for this field
74 as well. This paper is structured as follows: 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 it delineates the main research questions. Section 4 presents the findings of the study in two large subsections. Section 5 is the area of discussion of these findings, while in the last section, we conclude our study.

2. Background

OA and OS are considered as being movements for the restoration of scholarly communication values and the establishment of links with society and innovation (Suber, 2012). Research funding organizations have pushed towards the adoption of progressive policies and OS is often greeted as a beneficial set of practices for researchers (McKiernan et al., 2016). The authors of this study delineate the benefits of OS in an ecological system of openness, from Open Access to publications, to Open Data and then to Open Source, and conclude that these can take the form of increased attention to researchers' work, higher impact, more collaboration opportunities, etc. Berg and Niemeyer (2018) add other reasons why OS is required in engineering fields, such as achieving reproducibility, establishing priorities, encouraging trust, and so on. In the view of Frankenhuis and Nettle (2018), OS is beneficial because it emphasizes on quality, a concept that is controllable by communities themselves, instead of third parties, and on creativity, a concept that encourages scientific exploration instead of 'destination'. The Open Science Monitor (European Commission, 2019) observes 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 in order to understand how national dynamics are implemented. In its most recent version (SPARC Europe & Digital Curation Centre, 2019), the review concludes that European legislation on the use of public sector data has been a driver of developments in research data adoption, while national policies in place have been increased.

The ability to search in almost every scientific field a critical mass of information that someone can analyze and draw conclusions on, as well as the proliferation of data gathering or usage of 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 researchers of this field, according to a study by Rodrigues, Taga and Passos (2016). 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.

OA has various types, depending on the distribution channels, the financial aspects and the permits. The established concepts of OA have been color coded as Gold, Green, Bronze and lately Black. Gold is usually related with access to financially enabled publications in pre-existing venues, while Green relies mainly on academic or community driven infrastructures, either complimentary to Gold or self-sufficiently. Licensing ambiguity of Gold publications can lead to the Bronze state, whereas access through legally dubious means is related to the Black OA, which mostly affects legacy publication schemes. Björk et al. (2010) 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% of availability in formal and traditional publishing services and 12% for availability in repositories and other web services. The share of OA journals and their publications are growing at much faster rates than the rest of the literature and have now stabilized their production, availability and exposure (Laakso et al., 2011), while Björk, Shen & Laakso (2016) have explored the resilience of early publication venues that move between the lines of Gold and Diamond OA (another coloring for the free coverage of all costs, both for publication and access).

Archambault et al. (2014) provided a thorough, large scale review of the OA availability, which reaches up to 50% of the existing literature. The study attributes such high rates partially

to the backfiling of publications in repositories and websites and to the growth of Gold OA publications. A projection of the growth of OA has predicted that, in 2025, 70% of the readings of the scientific literature will be in OA, while the rate of these publications will climb from 30% to 45% (Piwowar, Priem, & Orr, 2019). The current 30% has been confirmed by a previous study of Piwowar et al. (2018) 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 (that is open publications but with limited licensing information). 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 (2014). 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 in conceptual ambiguity that stalls a unified comprehension of OA. A similar study in Catalonia (Rovira, Urbano, & Abadal, 2019) has reported an impressive 70% availability of at least one OA 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, which requires, as other national legislation, immediate public availability.

The commitment to the OA guidelines has been explored by researchers of the Library & Information Science domain, that carries a considerable weight of the OA advocacy. Grandbois and Beheshti (2014) have studied the open availability of LIS articles and found that 60% of these were 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 (2018)

has explored if researchers of OA “*practice what they preach*” and she found in her study that slightly over 25% of the articles are paywalled, which leaves an impressive 75% in various OA modes.

The bipolarity of OA and paywalled publishing has led to a discussion about the possible benefits of OA. The ways to measure ‘openness’ and the practices it envelops, as well as the ways capitalizing them for research assessment purposes, have turned to be a key interest for researchers (Levin, Leonelli, Weckowska, Castle, & Dupré, 2016). Recent initiatives have been undertaken for the development of ‘open’ profiles sketching (Chen & Olijhoek, 2016; Murphy & Jones, 2020), while others, such as the Open Science Career Assessment Matrix (Rentier, 2019), operate as check points that can inform research assessment exercises.

The ‘citation advantage’ of OA publications is a claim that has been supported by many of its advocates. It is a concept associated with two main investigations, one assessing the outcomes of citation production between paywalled and OA publications, and one exploring the causes, i.e. the reasons leading to any differences. Recent large scale studies (Wizdom, 2018) have exhibited significant differences in citation indicators, like Relative Citation Ratio (Hutchins, Yuan, Anderson, & Santangelo, 2016), that reach up to 130% between the citations of OA and paywalled articles, while others (Huang, Yue, Chen, Xu, & Li, 2019) have provided evidence that the OA status is not only responsible for increased citations, but, consequently, may benefit the Impact Factor of the medical journals that they are published in. While the increased citation performance of OA papers has been exhibited in numerous research studies (Breugelmans et al., 2018; Eysenbach, 2006; Hajjem, Harnad, & Gingras, 2006; Swan, 2010), Björk and Solomon (2012) supported the claim that there are similar citation rates between OA and paywalled works. Similarly, Craig, Plume, McVeigh, Pringle and Amin (2007) have articulated a critique towards the ‘advantage’ claim and argue that the main effect is not attributed to access to sources, but to their discovery. Other studies (Davis, Lewenstein, Simon, Booth, &

Connolly, 2008) have exhibited that OA can lead to usage advantage, but this does not secure a head start for increased citations. This has been challenged by Archambault, Côté, Struck and Voorons (2016) who believe that OA publications have shorter citation windows due to delays in Green OA availability, while Lewis (2016) has argued that ‘black OA’, or else pirate access, to paywalled publications is more likely to degrade the observed citation differences and sustain the current status of citation performance. According to a study by Evans and Reimer (2009), advantageous citation performances, even compared to free online and early publication, might result by a combination of traditional publishing services and OA status.

3. Research Questions & Methodology

In this study we explore the OA and OS literature, as published in a period of twenty years, from 1999 to 2018, and indexed in the bibliographic database Web of Science (WoS). While the two topics have differences in scope, we preferred to focus on both, as a representation of the two broadest subjects in 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 indices, namely SCI EXPANDED, SSCI, A&HCI and ESCI, for publications in the fields of OS and OA. We limited the scope of our query to the ‘Title’ field, used the terms ‘open access’ and ‘open science’ as the prime ones, 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 exact query can be found in Appendix 1. Tests with relevant terms, such as ‘open research’ and ‘open data’, were also performed, but we have opted for not using them, since they provided irrelevant results in the manner that these were not addressing explicitly the topics from the openness viewpoint. One of the reasons that we have opted for the WoS database was the versioning scheme that the database uses -in partnership with Our Research’s Unpaywall application (<https://unpaywall.org>)- for the access status of every publication. WoS assigns to each publication one or more labels of the main access types, namely Gold, Green and Bronze for all the aspects of OA types, and Paywalled for the toll type of access. To avoid overlapping we classified a publication to a single access type by first rank order.

The results were narrowed to those of the years 1999-2018 and the initial export was 3,042 records that -after further manual cleaning of papers that were not discussing openness- resulted in 2,823 records. These records were pre-processed to remove redundant columns and to check for syntactic consistency and error-prone characters. Frequent values in some columns, such as in the access type column, were coded to integers to enable accurate processing. In this stage, variations were grouped to large categories; an example being the ‘Gold Open Access’ option, which also included the ‘DOAJ Gold’ records. Computational 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

In this section, we explore the broad picture of the literature by presenting the statistic properties of major entities, namely sources, authors and publications, and the ways these have progressed over time (see Figure 1). The presentation focuses on the access status of these entities and investigates whether it can have an effect on the performance of publications or the profile of the authors.

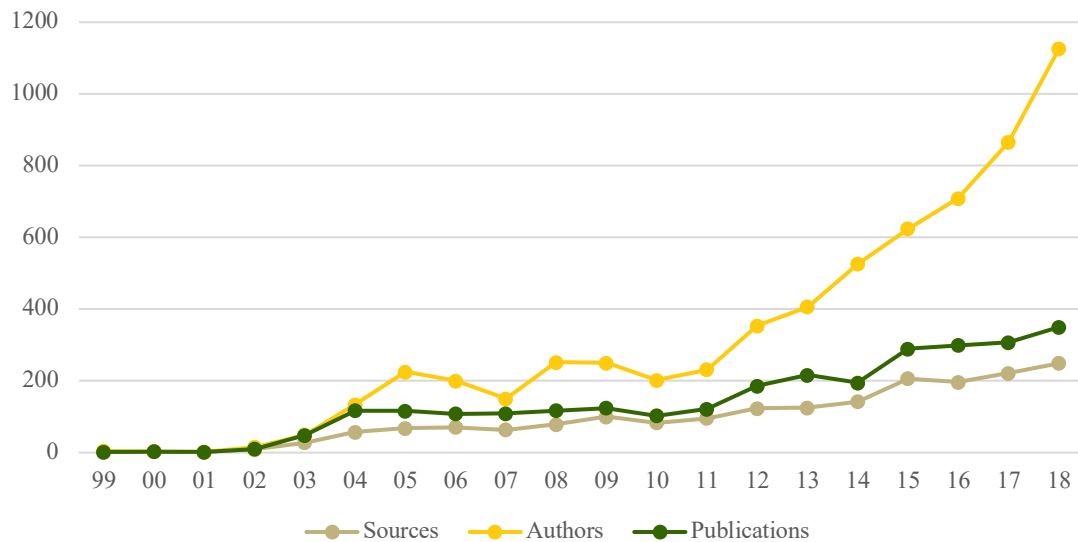


Figure 1. Timelines for sources, authors and publications

4.1.1. Sources

In the WoS nomenclature, ‘Sources’ is a term that describes publication venues, which, most of the times, corresponds to scientific journals. There were 1,183 distinct sources in our dataset, and by associating the access status of the publications to sources, we find out that 85 (6.54%) have allowed Green deposition in repositories, 161 (12.38%) were linked with the Bronze publishing mode, and 353 (27.15%) have published articles as Gold OA. The majority of the journals though (n=701, 53.92%) have published items in the toll access mode. Figure 2 presents the histogram of the number publications per source. The bins on the x-axis represent intervals of numbers of sources and the values on the y-axis (in logarithmic scale) corresponds to the number of publications, which have been published in the sources of each bin.

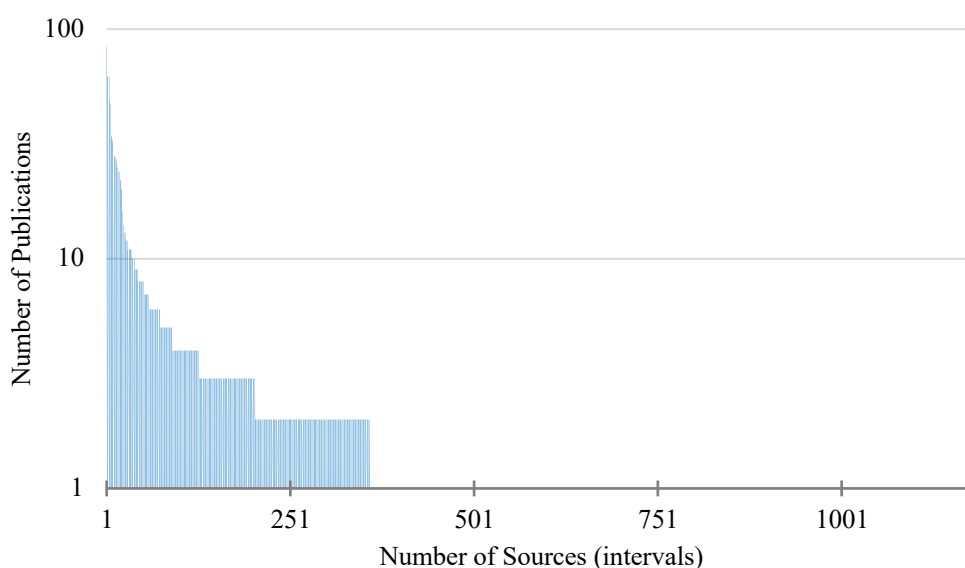


Figure 2. Histogram of publications per sources

Since 2012, the number of sources publishing items on the topics of OA/OS is consistently increasing. The last five years there was a 75.35% increase (from 142 sources in 2014 to 249 in 2018). The first four years of the period under question (1999-2003), the growth appears rather weak, but in the following years, from 2004 to 2011, it was moderate, which means that from 50 to 100 titles each year were publishing relevant papers. These sources are engaged into the discussion in various ways and not just in publishing research articles. Interestingly, more source titles have published editorial notes ($n=615$) than research articles ($n=515$), which is the prime type of contributions in the dataset. Reviews have been published in 92 sources, while the Editors of 79 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.

From a list of the 35 most frequent sources that have published on this topic more than ten articles each, we have manually checked through the JournalTOCs service (<http://www.journaltoCs.ac.uk>), which ones of those were entirely OA, entirely paywalled and hybrid sources (Figure 3). Thirteen sources, that publish 37.63% of these works, were hybrid, meaning that they provided a paid OA option to authors next to the costless, yet subscribed one. Eleven of

the sources are entirely OA, with two of them being ‘free’ (giving partial access), while, also, eleven of them were behind tolls. The OA sources have published 284 works (29.77%) and the paywalled 311 (32.6%). The greater part of this list belongs to the LIS discipline, a category followed by multidisciplinary (n=5) and medical science journals (n=4). Journals from chemistry and physics follow in the ranking, while the remaining two belong to biology and computer engineering.

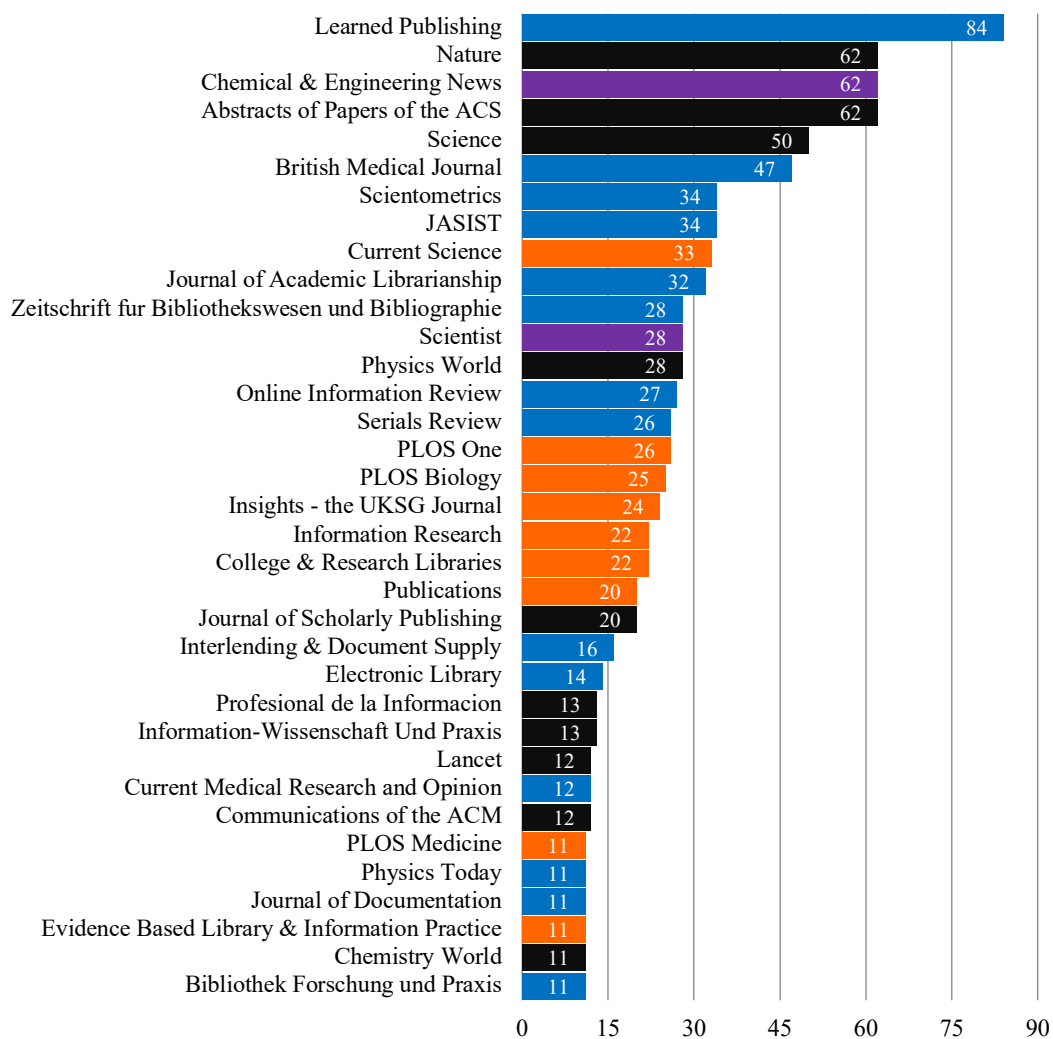


Figure 3. Top 35 sources and their access status (orange: Open Access; black: paywalled; blue: hybrid; purple: free)

4.1.2. Publications

In the 2,823 publications in our dataset, we have identified cases where two or more publications shared the same title. Such is the case of the title ‘Open Access’, which is common for 18 distinct publications.

These publications were grouped in seven document types, namely article, correction, editorial, letter, news, review (book, conference, etc.) and unlabeled, with the latter grouping smaller representations of other document types. Article is the main type in our data (n=1,166, 41.3%), followed by editorial notes (n=906, 32.09%). At the same time, there was a strong representation of letters to the Editors (n=218), while the popularity of OA/OS issues was also reflected on a significant amount of news items (n=223). Until 2004, the main types of contribution appear balanced. Letters to the Editors are present after 2000 and onwards, while research articles have a significant growth after 2015 (84.21% from 2014). However, in three years, 2004, 2005 and 2012, the number of research papers is lower than that of editorial notes. The latter grow substantially from 2012 and onwards and in 2015 hold an impressive record of 100 publications of this type.

Only a small number of these publications has featured a funding note (n=340, 12.04%), which implies that the majority have not been funded (n=2,483, 87.96%). The first funding note has been traced in 2003 and the phenomenon is very rare in the first five years of the two decades (1999-2004 and 2010-2015). In general, the majority of these notes appear in research articles (269, 9.53% of all publications), but at the same time most of the articles (n=897, 31.77% of total) together with editorial notes (n=871, 30.85% of total) do not seem related to some sort of funding. From the ones with funding notes, 43.53% (n=148) are in Gold OA and a similar 42.06% (n=143) in paywalled journals.

In bibliographic databases, such as Web of Science, terms that accompany the description of a publication help its indexing and classification. There are two sources for these terms:

descriptors, which are keywords that the authors themselves have chosen to describe their publications, and identifiers, or else Keywords Plus®, which are controlled terms that WoS assigns to publications. In both cases, there is a percentage close to 72% of having no term assigned, see 2,023 with no descriptors and 2,040 with no identifiers. The remainder publications (n=783) featured -in overall- 1,650 identifiers, while 800 publications featured 2,293 descriptors. Terms of either source are mostly present in Gold OA and paywalled publications at a rate of one third in each type and concentrated mainly in articles and editorial notes.

In our data, almost 57.6% of the publications were behind subscription walls (n=1,626, 57.60%), while Gold OA represented approximately 27% (n=758, 26.85%), followed by Bronze (n=311, 11.02%) and Green (n=128, 4.53%). Paywalled publications have significantly increased after 2012 (50.72% since the previous year, 2011) and they remained to these levels with a tendency to further increase. A year of similar growth for Gold OA publications can be traced in 2014. Regarding the other two categories, Bronze publications have been present in all years since 2003, and Green, although in lower absolute numbers, has shown signs of growth after 2015.

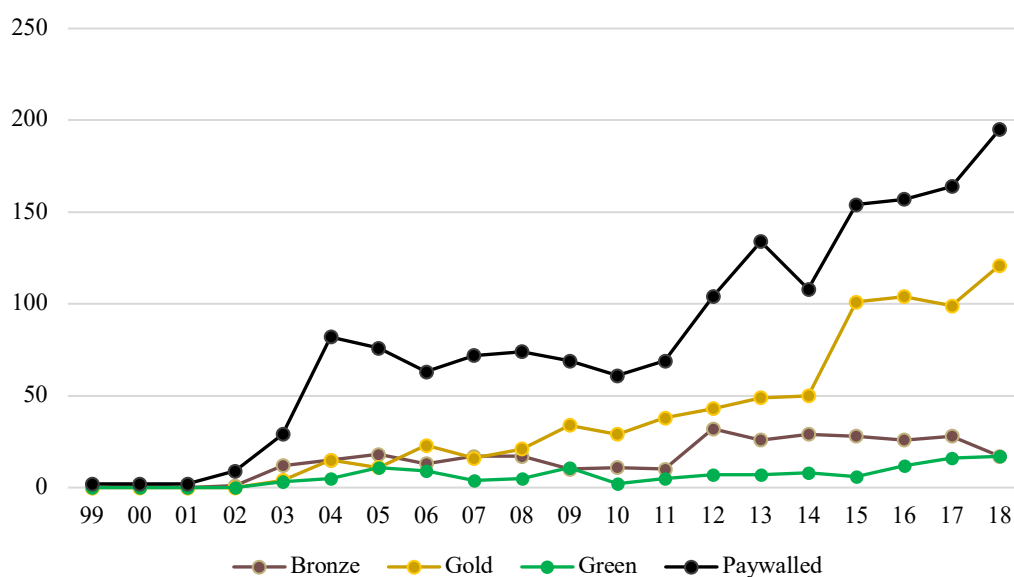


Figure 4. Timeline of publications by access status

There are 1,370 publications (48.53%) that have never been cited, and 385 that have been cited only once (13.64%). At the same time, there were 80 papers that have been cited five times (2.83%), 21 papers cited 10 times (0.74) and nine papers that have been cited 20 times (0.32%). Figure 5 presents the distribution of publications per access type according to the number of citations (up to 20).

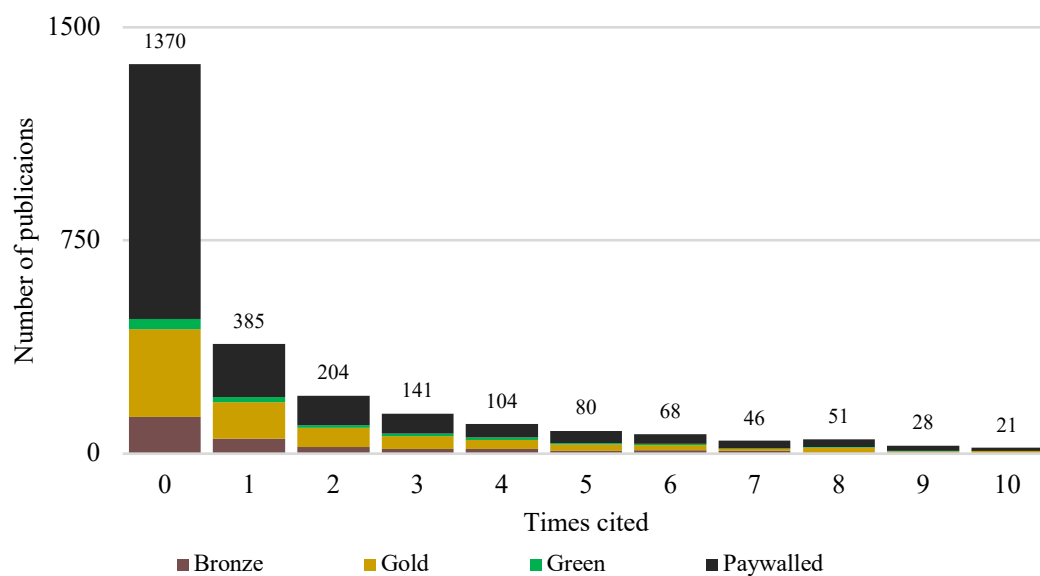


Figure 5. Number of publications and number of citations per access type.

The most cited paper, an annual report of an OA database (Sandelin, Alkema, Engström, Wasserman, & Lenhard, 2004), has received 855 citations and it was responsible for the high number of citations in 2004, the year of its publication. Similar is the case of the year 2008, when there was a new version of this report that accounted for 467 citations (Bryne et al., 2007), while both years have the highest average of times cited, i.e. 14.18 and 14.01 respectively. There were also three years -2006, 2009 and 2010- when the average citation number exceeded ten citations per publication, see 10.20, 10.15 and 10.70, respectively.

In our study we examined whether differences can be traced in the citation performance of each access type. Figure 6 shows that there was an evidence of ‘citation advantage’ for all OA

publications and most notably for those published in Gold road. Gold publications, which represent 26.85% of the overall number, have received 7,816 citations that correspond to 43.51% of overall citations. Similarly, Bronze and Green represent 7.94% and 12.17% of overall citations, while the citations produced by paywalled publications, which corresponded to 57.6%, are 36.38%. Thus, the average number of citations per each access type was 4.02 for paywalled, 4.59 for Bronze, 10.31 for Gold and 17.08 for Green publications. It must be noted that the number of Green publications was enforced by that one publication that has received 855 citations alone.

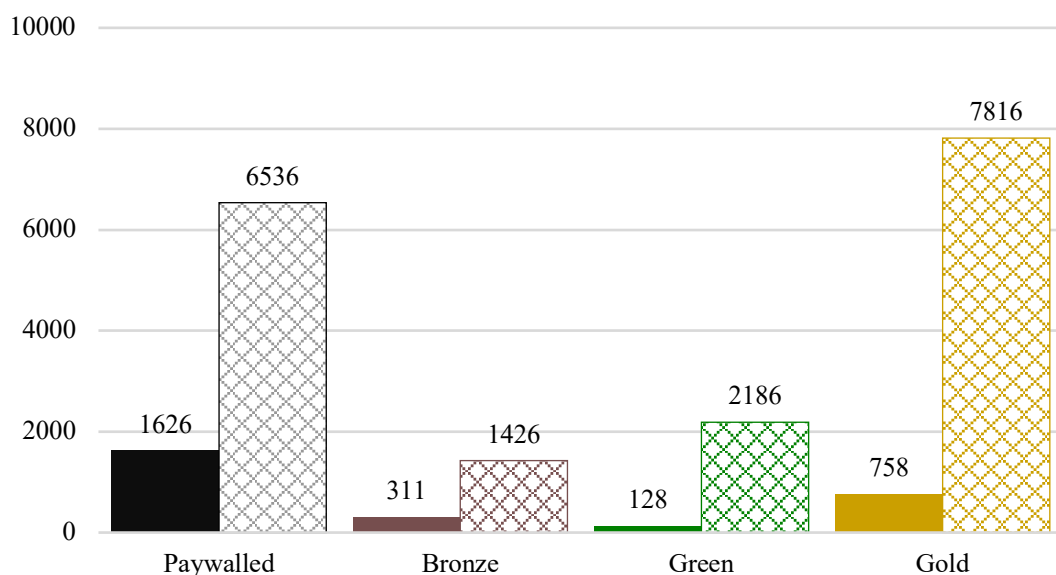


Figure 6. Number of publications (solid bars) and number of times cited (pattern bars) per access type.

4.1.3. Authors

The 2,823 publications of our data have been authored by 2,342 authoring entities that range from 1 to 138 co-authors in a work. When split and deduplicated, these entities resulted in 5,542 individual authors. Figure 7 presents the histogram of publications per author.

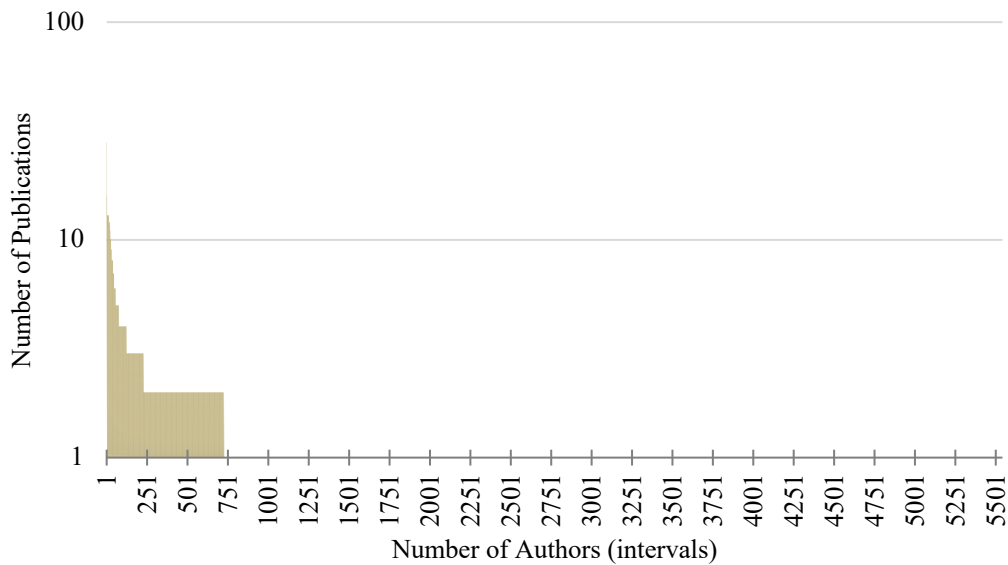


Figure 7. Histogram of publications per author (y-axis in logarithmic scale).

From 2005 to 2011, the number of individual authors has been fluctuating between 150 and 250. This number has 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 were the highest during that year. Indeed, the average number of authors per publication in this year was 3.22, while during the last five years it was steadily above the average number of two authors per publication.

It is notable that in 2014 there were more authors publishing Gold OA contributions, than paywalled; a year in which the average number of authors in this access type was the second highest, 4.6 (after 2008, which was 4.71). For an entire decade, from 2007 to 2016, there are more authors in Bronze than in Green OA publications. However, the last three years, the average number of authors who published their work in Green OA was increasing, see 5.17 in 2016, 6.19 in 2017 and 12.35 in 2018. This seems to have affected the average number of authors for self-archived publications, which is 5.47, while the average for paywalled is 2.10, for Gold is 3.14, and for Bronze OA publications is 1.85. The mean number of authors for

articles is 3.19 and a similar mean is for corrections is 3.71, whereas on the other side the mean number for news and letters is 1.05 and 1.98, respectively.

Based on the count of the access types of publications an openness score of publications was produced, as the ratio of all OA publications (of any kind) to all publications of an author in the dataset. The inspection of the corresponding histogram showed that the majority of authors stood on the two ends of the scale, either open, or paywalled. This can be explained by the low number of contributions by each author, which could belong to either the closed, or to any of the open access types. In order to highlight notable performances, this score was produced for a subset of authors that were selected with the criterion of having published more than ten contributions in the field. Figure 8 presents the list of selection of the top 22 authors with their respective proportion rate, excluding though the ‘Anonymous’ author, which alone has featured in 117 publications.

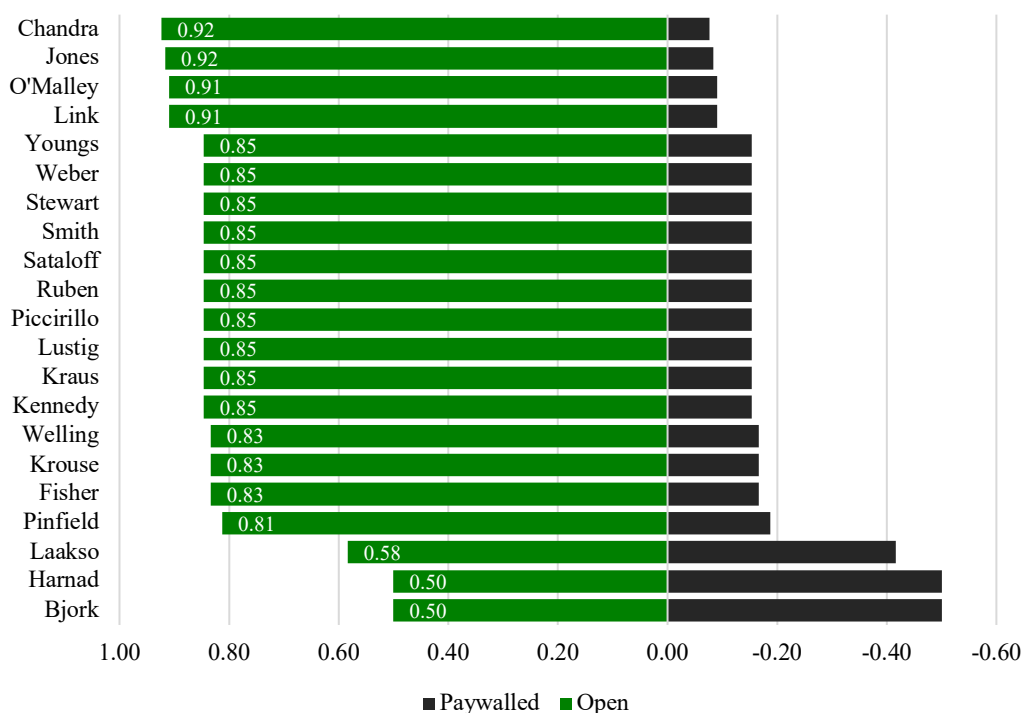


Figure 8. Openness scores for top 22 authors.

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 have split the results to (i) practice variations, analyzing the publication practices of particular authors and sources that have been qualified based on their activity, (ii) population behavior, concentrating on the consistency of authors, both in general and per access type, and the expansion rate of the field, and (iii) publication loyalty, exploring microscopically what levels of consistency could be traced in publication patterns.

4.2.1. Practice variations

Variations for top authors

The earliest record by the group of the top 22 authors, as defined in section 4.1.3, could be found in 2004, but the main part of their activity was recorded in the last three years. There were 294 publications by these authors that, in the wide selection of journals, have been published in 36 unique sources. Checking in the WoS categories for the disciplines of these sources revealed that most of them belonged in the medical sciences (15 sources) and LIS (six sources), while 31 of these have been ranked in the Journal Citation Index. The number of sources, following the trend line of the publications, was stably low during most of the years, showing however a rapid increase in the last three years.

In the majority of these years, the top 22 authors published their contributions in paywalled journals in almost every year. The consistency of publication was rather disturbed in the last three years, and this was evident also in the access types of publications of these authors. The boost in 2018 has resulted in a large number of authors in Green OA publications and, despite the fact that the authors with Bronze OA were few in numbers, the other two categories of OA have displayed high scores, even higher than paywalled publications. This small anomaly was attributed to 13 publications by an authoring entity of 23 authors, exhibiting varied composition. This was the case of a common note by the Editors of 13 publications in the broad field of

otolaryngology that have commonly expressed their concerns under the title “*Open access: Is there a predator at the door?*”

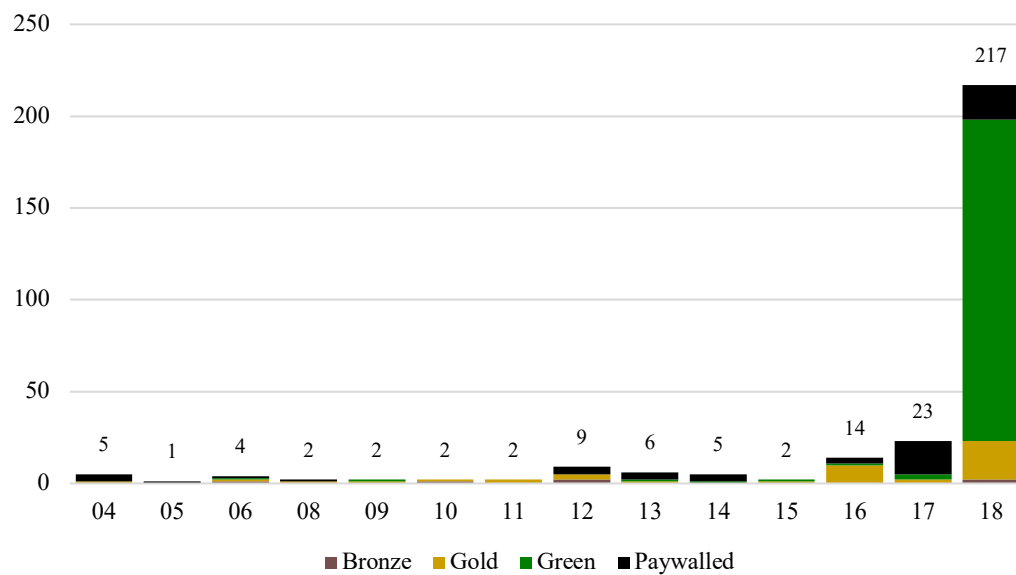


Figure 9. Variations of top authors’ publications per access type.

The variations of document types of these authors during the entire period reflected a diverse landscape of contributions. The only year that there was invariability was the last one, 2018, which was affected by the case of combined editorial notes. This type of documents was followed by research articles, but it should be noted that in 2017 there was a relatively high number of letters to the Editor by these authors, that even surpassed the number of articles.

Variations in top sources

The exploration of variations in particular sources was focused on the selection of the 35 journals that have published ten or more contributions in the field. These 35 journals have hosted the publication of 951 papers by 1,281 unique authors, with the densest representation in years 2013 and 2017, 88 and 86, respectively. Fourteen out of these sources were also in the list of 36 journals that the 22 most active authors have published in, while from their 1,281 authors, there were four who belonged to this group.

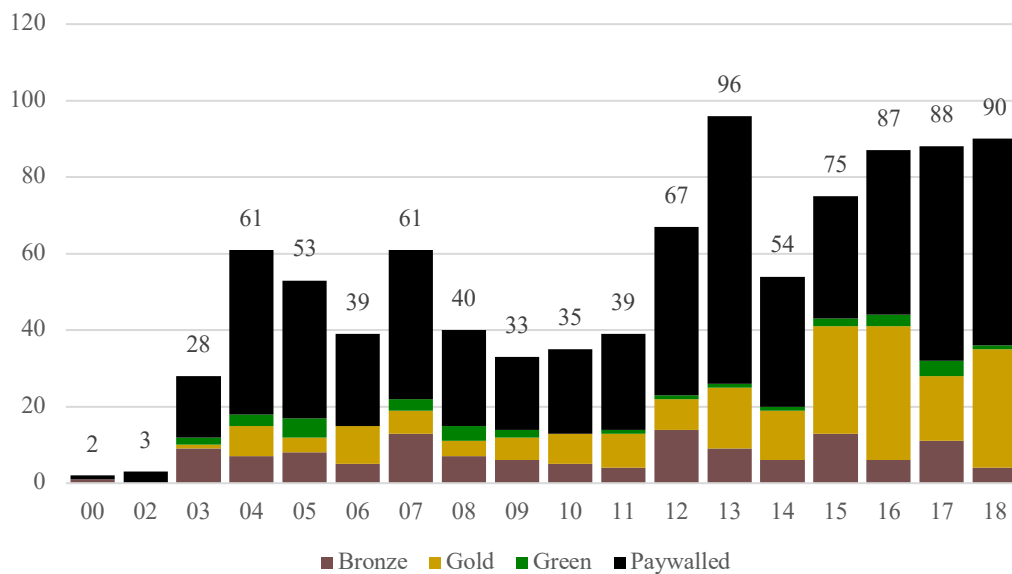


Figure 10. Variations of top sources' publications per access type.

Figure 9 shows the access type of publications in the top journals during the entire period. As most of these 35 journals were either subscription or hybrid ones, it was reasonable to meet paywalled publications in every year. The figures of paywalled publications were the highest ones yearly, being challenged only by the Gold OA publications in 2015, while Green OA publishing has been present every year from 2011 onward.

Research articles and news items were 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. A considerable number of editorials have been published in Nature in the last two years.

4.2.2. Population behavior

Every research field is developed over time and authors are the driving force as through their publications they help to shape its understanding and direction. Thus, we examined the presence

of authors through time, in order to understand if scholars of this field were consistently interested in it. To do so, we gave authors four labels, depending on their behavior: (a) first-time authors were those who first appeared in a publication in a year and have appeared again in any later year of our dataset, (b) last-time authors were the ones that have appeared in any past year, but they appeared for the last time in a specific year of our dataset, (c) one-off authors appeared in a publication once, but they did not appear again and (d) repeating authors were the ones that appeared in publications in consequent years. Figure 11 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 62.7%. On the other side, the category that showed a persistence in research publications of this topic, the repeating authors, have quite low rates, which has a minimum of 2.3% (2017) and a maximum of 9.9% in 2010.

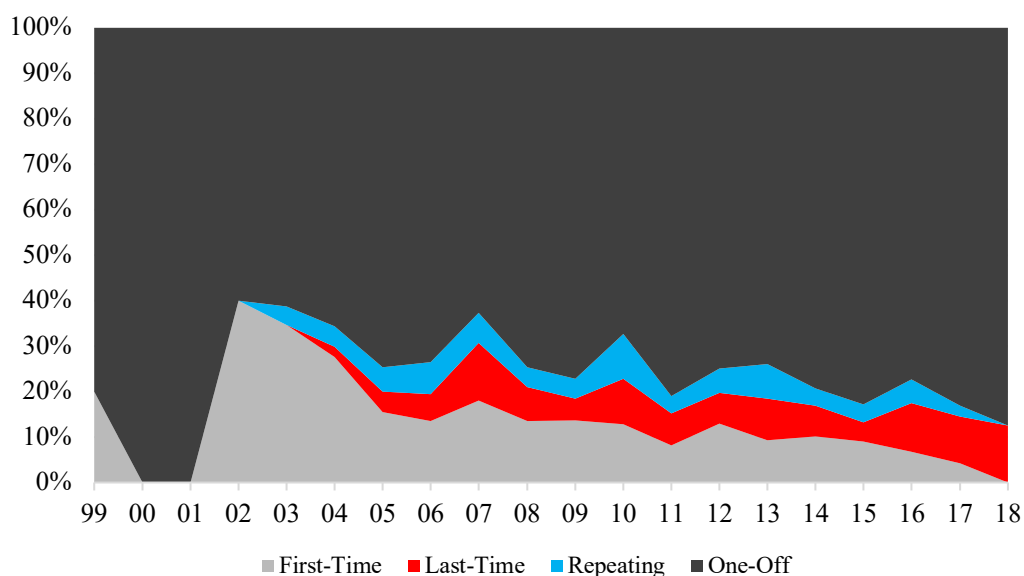


Figure 11. Population behavior. In 1999 there can be no last time and repeating authors, while in 2018 can be no first time and repeating authors.

Many of the repeating authors have published a considerable amount of Bronze OA contributions in 2013 (10.5%), while a similar consistency by repeating authors was recorded for Gold OA contributions in 2010 (11.6%). The rate of repeating authors in paywalled articles has

been quite low (from 2.5% to 5.9%), while the number of first time authors has been decreased in the last five years.

This information allowed the development of an expansion rate in order to estimate the streams of new authors during time, which was calculated with the count of the one-off and the first time authors to all authors in a year. According to this rate, 2003, 2004 and 2015 were the three years that most new authors entered the field with their publications, with a rate of 0.95, 0.93 and 0.91, respectively. In 2009, paywalled contributions expanded by 0.94 and similarly in 2011 the Gold OA publications expanded by 0.96. As far as Bronze OA concerns, the year with the highest expansion rate is 2015, with 0.96, and the last year in the period, 2018, has been the year with more new contributions (0.97).

4.2.3. Publication loyalty

In section 4.2.2, we have explored the publication behavior of authors and how it shaped the contributions in the field. We have zoomed in the publication history of authors to further explore the consistency of their presence, or else to identify a continuity pattern. We first measured the number of consequent years of publication in the same source and found that two authors, M. Banks and A. Widener, have published for four consequent years in the same journal, Physics World and Chemical & Engineering News respectively, followed by a group of twelve other authors with three publications in subsequent years. On the other hand, there were also notable periods of absence from publishing in the same source, meaning that we calculated the number of years between two publishing instances of an author. In this case, two authors, D. Butler and R. Anderson, had a considerable amount of years before returning to publish in a source in which they had published before, the former after nine years in Nature and the latter after eight years in Learned Publishing. These two authors were followed by a group of seven authors with seven years of absence.

Similarly, we measured the cases of journals that had published works of the same authors for a number of consequent years. There were eleven journals that have hosted three publications in a row of the same author, while Physics World and Chemical Engineering Letters have welcomed contributions for four consequent years by the same author. The reverse pattern, this of lapse, for sources had similar periodization. The majority of the journals, six, have seen the same authors returning to publish after eight years, while there were titles, such as Nature, PLOS Medicine and Learned Publishing, that have eight or more years of hosting publications of previously appeared authors.

Certain journals, such as Nature and Learned Publishing, seem to publish research classified as Bronze for a large period, for 15 and 13 years respectively, while hybrid journals, such as Serials Review and JASIST, publish -for ten and nine years respectively- in toll-access mode. Two journals from the PLOS family, Biology and One, together with Current Science publish for more than eight years in row items in Gold OA. While in the case of Serials Review the publications that appear every year in a period of eight years are Editorial Notes, for 15 years in Learned Publishing and for eight continuous years in JASIST and PLOS One the most preferred type is research articles. On the contrary, sporadic publications on OA/OS are hosted in less frequent journals, such as Swiss Medical Weekly and Plant Methods, which exhibit a period of 11 and ten years since publishing again Gold OA articles, as well as in Anesthesia and Analgesia, which has published after ten years in paywalled mode. Similarly, the type of contribution that has been absent for most years in sources like British Journal of Nutrition, American Journal of Physics, American Scientist and Bioscience is Editorial notes for more than eight years.

5. Discussion

508 The field of OS/OA is rapidly evolving in the last five years, although substantial growth has
509 been recorded since the start of the second decade. This is in accordance with other studies that
510 focus on the growth of OA documents in general (Kurata, Morioka, Yokoi, & Matsubayashi,
511 2013; Liu & Li, 2018), but it also holds similarity to studies of other research topics, like re-
512 search data management (Zhang & Eichmann-Kalwara, 2019). This shows that the tectonic
513 plates of scholarly communication are on the move, shaping a different landscape through in-
514 tense communication. The reasons, of course, can be multifaceted and it can be attributed to the
515 increased interest in the field, as well as to commentaries on developments, such as the NHS
516 policy in 2008 and the various subscription cancellations in US and EU Universities from 2017
517 and onwards. Even though there are differences in the OA representation, our findings on the
518 growth in OA/OS publications are similar to Schultz's, who has indicated 2013 as a strong
519 turning point for their increase. In our study, the majority of publications is still in paywalled
520 mode, but the OA options gradually become more competitive. Despite also the notable citation
521 performance differences that have been found in favor of OA in this study, this literature is still
522 being published in subscription venues. The Gold OA road represents an option that the authors
523 of more than one quarter of the publications have preferred, especially in the last five years.
524 However, this contrasts with other studies, like Gargouri, Larivière, Gingras, Carr and Harnad
525 (2012), who have evidence of increased availability of Green OA publications over Gold ones.
526 In the present study, the Green OA road has exhibited publications with large compositions of
527 authors, especially during the last three years. The size of Bronze OA publications is also alarm-
528 ing and it has concerned other research teams, such as Piwowar et al. (2018), who identify large
529 concentrations of this phenomenon in high-profile publishers. It is again difficult to assume
530 what has lead -unintentionally or not- publishers and authors to this practice, but it is most
531 certain that the *“fleeting, unfair, and asymmetrical nature of power that this type of access*

embodies”, as stated by Eamon (Costello, 2019), obscures the status of OA publications and hampers one of the key notions which is reusability.

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, one can understand that the field has a greater part of authors with limited connection. The rate of authors who publish consistently in the field is quite low, ranging from five to ten percent and these ‘returning’ authors prefer the Gold OA 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 OA road also exhibits a consistent rate of new authors every year, in contrast to all other access types. 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 participating in the emerging dialogue. While LIS journals have a strong interest in this topic, this discourse is not their exclusive right. Multidisciplinary journals, as well as journals from medicine, chemistry and physics, are also platforms for communication of findings and opinions. The sample of the most active sources in our dataset revealed that the majority were either subscribed, or hybrid venues, while almost one third was entirely OA. Thus, the road to full OA seems long and, as recent author surveys have shown (Nature Publishing Group, 2015; Taylor and Francis, 2019), the prestige of a journal or its esteem in the community are more determining reasons to select a publication venue than its access status. Although one cannot firmly identify the reasons for selecting a journal, prestige, as substantiated by the use of IF, can be linked with the selection of the top publication venues in our study, as 31 out of the 36 feature this indicator.

Similarly, in the absence of explanatory approaches, it is risky to speculate why one author prefers or not to publish openly; it can be any institutional, cultural, financial and technical

reasons at any time during these twenty years, which of course is dense time. Therefore, the openness score we presented cannot explain the reasons why an author publishes in a journal, but it gives an indication of the consistency of the field. The top authors in our dataset seem to have a very consistent behaviour and one could claim that they actually ‘practice what they preach’.

With a sizable dataset and microscopic exploration of the behaviors, we were able to reveal practices of interest. Documents that are published in subsequent versions and are addressed 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 in citation performance. It is difficult though to draw conclusions about the reasons for these differences, as the observed performance has been measured at the end of a long period, during which many conditions, including selection bias or quality bias, might be enacted. However, since the dataset is not referring to specific sources or specific institutions and taking into account the low numbers of Green OA publications, mandated OA does not seem to be one of the reasons. On the contrary, the combination of traditional journal publishing with self-selected self-archiving in Green OA infrastructures seems to be responsible for increased citation performance.

Although the citations of the last two years can be excluded, 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 are not regularly receiving many citations. The large number of editorial notes, especially the last few years, shows that there is an escalation of intensity in the dialogue about OA/OS, in which concerned Editors wish to take part. The average number of citations per access type shows that Gold and Green OA enjoy more citations than Bronze and paywalled. When we omitted the calculation of the extreme citation

score of a self-archived article, we witnessed that the average numbers of Gold and Green are very close (10.31 for Gold and 10.40 for Green), while the other two remain below five citations per publication.

Inevitably, our study has certain limitations including the use of a specific bibliographic database, with a highly selective policy and Anglo-Saxonic orientation, to retrieve our data. Much of the discourse has been published in other means than apart from scientific journals, such as trade journals, reports, working documents and statements, while not all journals from all countries are listed in this database. Therefore, together with technical conditions, such as the specific query, our data do not claim comprehensive representation. Further analyses could be performed if indexing services had acted to lower the scores of null values in identifiers and descriptors or had labelled the type of contributions.

6. Conclusions

OS and OA are two topics that have concerned intensely researchers from all scientific disciplines and focus on fundamental changes in scholarly communication that the current system 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 to reach wider audiences. Furthermore, a significant part of the literature is under obscure licenses and thus omissions or intentional acts that result to states, such as Bronze OA, should be decommissioned. Finally, in order to make OA a reality new authors should populate more stably the field, as the current research has shown that only a small number remains active. It is important that every part of the discourse for OA/OS remains open and accessible to anyone who wishes to understand and to contribute.

References

- Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). *Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013*. Retrieved from <http://scimetric.com/en/publications/reports/proportion-of-open-access-papers-published-in-peer-reviewed-journals-at-the>
- Archambault, Éric, Côté, G., Struck, B., & Voorons, M. (2016). *Research impact of Paywalled versus Open Access papers*. Retrieved from <https://www.1science.com/1numbr/>
- Berg, D. R., & Niemeyer, K. E. (2018). The case for openness in engineering research [version 2; peer review: 3 approved, 1 approved with reservations]. *F1000Research*, 7(501). <https://doi.org/10.12688/f1000research.14593.2>
- Björk, B.-C., Shen, C., & Laakso, M. (2016). A longitudinal study of independent scholar-published open access journals. *PeerJ*, 4, e1990. <https://doi.org/10.7717/peerj.1990>
- Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, 10(1), 73. <https://doi.org/10.1186/1741-7015-10-73>
- Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open Access to the scientific journal literature: situation 2009. *PLOS ONE*, 5(6), e11273. Retrieved from <https://doi.org/10.1371/journal.pone.0011273>
- Breugelmans, J. G., Roberge, G., Tippet, C., Durning, M., Struck, D. B., & Makanga, M. M. (2018). Scientific impact increases when researchers publish in open access and international collaboration: a bibliometric analysis on poverty-related disease papers. *PLOS ONE*, 13(9), e0203156. Retrieved from <https://doi.org/10.1371/journal.pone.0203156>
- Bryne, J. C., Valen, E., Tang, M.-H. E., Marstrand, T., Winther, O., da Piedade, I., ... Sandelin, A. (2007). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Research*, 36(suppl_1),

632 D102–D106. <https://doi.org/10.1093/nar/gkm955>

633 Chen, X., & Olijhoek, T. (2016). Measuring the degrees of openness of scholarly journals
 634 with the Open Access Spectrum (OAS) evaluation tool. *Serials Review*, 42(2), 108–115.
 635 <https://doi.org/10.1080/00987913.2016.1182672>

636 Costello, E. (2019). Bronze, free, or fourrée: an open access commentary. *Sci Ed*, 6(1), 69–72.
 637 <https://doi.org/10.6087/kcse.157>

638 Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access
 639 articles have greater citation impact? A critical review of the literature. *Journal of In-*
 640 *formetrics*, 1(3), 239–248. <https://doi.org/https://doi.org/10.1016/j.joi.2007.04.001>

641 Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008).
 642 Open Access publishing, article downloads, and citations: randomised controlled trial.
 643 *BMJ*, 337, a568. <https://doi.org/10.1136/bmj.a568>

644 European Commission. (2019). Open science monitor. Retrieved from [https://ec.eu-](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)
 645 [ropa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-pol-](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)
 646 [icy/open-science/open-science-monitor_en](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en)

647 Evans, J. A., & Reimer, J. (2009). Open Access and global participation in science. *Science*,
 648 323(5917), 1025 LP – 1025. <https://doi.org/10.1126/science.1154562>

649 Eysenbach, G. (2006). Citation advantage of Open Access articles. *PLOS Biology*, 4(5), e157.
 650 Retrieved from <https://doi.org/10.1371/journal.pbio.0040157>

651 Fathli, M., Lundén, T., & Sjögarde, P. (2014). The share of Open Access in Sweden 2011 -
 652 Analyzing the OA outcome from Swedish Universities. *ScieCom Info*, 10(2). Retrieved
 653 from <https://journals.lub.lu.se/sciecominfo/article/view/11645>

654 Frankenhuis, W. E., & Nettle, D. (2018). Open Science is liberating and can foster creativity.
 655 *Perspectives on Psychological Science*, 13(4), 439–447.
 656 <https://doi.org/10.1177/1745691618767878>

657 Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and Gold Open
658 Access percentages and growth, by discipline. In Éric Archambault, Y. Gingras, & V.
659 Larivière (Eds.), *Proceedings of 17th International Conference on Science and Technol-*
660 *ogy Indicators, Montréal: Science-Metrix and OST* (pp. 58–70). Montréal.

661 Grandbois, J., & Beheshti, J. (2014). A bibliometric study of scholarly articles published by
662 library and information science authors about open access. *Information Research*, 19(4).
663 Retrieved from <http://informationr.net/ir/19-4/paper648.html>

664 Gutknecht, C. [@chgutknecht]. (2019). A report about the progress of #openaccess agree-
665 ments in Switzerland, which is closed access? Seriously??
666 <https://doi.org/10.1515/abitech-2019-3002>. Retrieved from [https://twit-](https://twitter.com/chgutknecht/status/1170028973935529984)
667 [ter.com/chgutknecht/status/1170028973935529984](https://twitter.com/chgutknecht/status/1170028973935529984)

668 Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-year cross-disciplinary comparison of the
669 growth of Open Access and how it increases research citation impact. *ArXiv*. Retrieved
670 from <http://arxiv.org/abs/cs/0606079>

671 Huang, C., Yue, X., Chen, J., Xu, W., & Li, J. (2019). The effect of “open access” on journal
672 impact factors: a causal analysis of medical journals. *Physica A: Statistical Mechanics*
673 *and Its Applications*, 533, 122043.
674 <https://doi.org/https://doi.org/10.1016/j.physa.2019.122043>

675 Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative Citation Ra-
676 tio (RCR): a new metric that uses citation rates to measure influence at the article level.
677 *PLOS Biology*, 14(9), e1002541. Retrieved from <https://doi.org/10.1371/jour->
678 [nal.pbio.1002541](https://doi.org/10.1371/journal.pbio.1002541)

679 Kurata, K., Morioka, T., Yokoi, K., & Matsubayashi, M. (2013). Remarkable growth of Open
680 Access in the biomedical field: analysis of PubMed articles from 2006 to 2010. *PLOS*
681 *ONE*, 8(5), e60925. Retrieved from <https://doi.org/10.1371/journal.pone.0060925>

682 Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The
683 development of Open Access journal publishing from 1993 to 2009. *PLOS ONE*, 6(6),
684 e20961. Retrieved from <https://doi.org/10.1371/journal.pone.0020961>

685 LERU. (2018). Open Science and its role in universities: a roadmap for cultural change. Re-
686 trieved from [https://www.leru.org/publications/open-science-and-its-role-in-universities-](https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change)
687 [a-roadmap-for-cultural-change](https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change)

688 Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How do scientists de-
689 fine openness? Exploring the relationship between Open Science policies and research
690 practice. *Bulletin of Science, Technology & Society*, 36(2), 128–141.
691 <https://doi.org/10.1177/0270467616668760>

692 Lewis, D. W. (2016). Will Sci-Hub kill the Open Access citation advantage and (at least for
693 now) save toll access journals? Retrieved from [https://scholarworks.iupui.edu/han-](https://scholarworks.iupui.edu/handle/1805/11161)
694 [dle/1805/11161](https://scholarworks.iupui.edu/handle/1805/11161)

695 LIBER. (2018). LIBER Open Science Roadmap. Retrieved from [http://doi.org/10.5281/ze-](http://doi.org/10.5281/zenodo.1303002)
696 [nodo.1303002](http://doi.org/10.5281/zenodo.1303002)

697 Liu, W., & Li, Y. (2018). Open access publications in sciences and social sciences: a compar-
698 ative analysis. *Learned Publishing*, 31(2), 107–119. <https://doi.org/10.1002/leap.1114>

699 McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... Yarkoni, T.
700 (2016). How open science helps researchers succeed. *ELife*, 5, e16800.
701 <https://doi.org/10.7554/eLife.16800>

702 Murphy, F., & Jones, P. (2020). *Openness profile: defining the concepts*.
703 <https://doi.org/10.5281/ZENODO.3607579>

704 Nature Publishing Group. (2015, September 23). Author Insights 2015 survey.
705 <https://doi.org/10.6084/m9.figshare.1425362.v7>

706 Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., ... Haustein,

- S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- Piwowar, H., Priem, J., & Orr, R. (2019). The future of OA: a large-scale analysis projecting Open Access publication and readership. *BioRxiv*, 795310. <https://doi.org/10.1101/795310>
- Priem, J. [@jasonpriem]. (2018). Wow. Did not realize that this entire special issue on #open-access is...paywalled. Can read whole thing for low price of \$300. Words fail me. Retrieved from <https://twitter.com/jasonpriem/status/948990852650676224>
- Rentier, B. (2019). *Open Science, the challenge of transparency*. Académie éditions. Retrieved from <https://academie-editions.be/accueil/369-open-science-the-challenge-of-transparency.html>
- Rodrigues, S. R., Taga, V., & Passos, F. M. (2016). Research articles about Open Access indexed by Scopus: a content analysis. *Publications*. <https://doi.org/10.3390/publications4040031>
- Rovira, A., Urbano, C., & Abadal, E. (2019). Open access availability of Catalonia research output: case analysis of the CERCA institution, 2011-2015. *PLOS ONE*, 14(5), e0216597. Retrieved from <https://doi.org/10.1371/journal.pone.0216597>
- Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research*, 32(suppl_1), D91–D94. <https://doi.org/10.1093/nar/gkh012>
- Schultz, T. (2018). *Practicing what you preach: evaluating access of Open Access research*. <https://doi.org/10.31229/osf.io/ybdr8>
- SPARC Europe, & Digital Curation Centre. (2019). An analysis of Open Science policies in Europe v4. <https://doi.org/http://doi.org/10.5281/zenodo.3379705>

731 Suber, P. (2012). *Open Access*. Cambridge, MA: MIT Press. Retrieved from [https://mit-](https://mit-press.mit.edu/books/open-access)
732 [press.mit.edu/books/open-access](https://mit-press.mit.edu/books/open-access)

733 Swan, A. (2010). *The Open Access citation advantage: studies and results to date*. Retrieved
734 from <https://eprints.soton.ac.uk/268516/>

735 Taylor and Francis. (2019). Taylor & Francis researcher survey 2019: researchers' views on
736 the present and future of scholarly communication. Retrieved from [https://authorserv-](https://authorservices.taylorandfrancis.com/researcher-survey-2019/)
737 [ices.taylorandfrancis.com/researcher-survey-2019/](https://authorservices.taylorandfrancis.com/researcher-survey-2019/)

738 Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H.
739 J. (2016). The academic, economic and societal impacts of Open Access: an evidence-
740 based review [version 3; peer review: 4 approved, 1 approved with reservations].
741 *F1000Research*, 5(632). <https://doi.org/10.12688/f1000research.8460.3>

742 Wizdom. (2018). Open Access - Global trends and rankings. Retrieved from
743 <https://www.wizdom.ai/dashboards/open-access>

744 Zhang, L., & Eichmann-Kalwara, N. (2019). Mapping the scholarly literature found in Scopus
745 on "Research Data Management": a bibliometric and data visualization approach. *Jour-*
746 *nal of Librarianship and Scholarly Communication*, 7(1), 0–19.
747 <https://doi.org/http://doi.org/10.7710/2162-3309.2266>

748

749 **Appendix 1. Query**

750 We searched in all Clarivate Analytics' Web of Science indexes, namely SCI-EXPANDED,
751 SSCI, A&HCI and ESCI for the following query:

752 ("Open Science") OR TITLE: ("Open Access") NOT TOPIC: (endoscop*) NOT
753 TOPIC: (fish*) NOT TOPIC: (enteroscop*) NOT TOPIC: (schedul*) NOT TOPIC:
754 (land) NOT TOPIC: (patien*) NOT TOPIC: (earth) NOT TOPIC: (DEM) NOT
755 TOPIC: (cancer) NOT TOPIC: (bowel) NOT TOPIC: (cell*) NOT TOPIC: (X-Ray)
756 NOT TOPIC: (densit*) NOT TOPIC: (NMR) NOT TOPIC: (gastroscop*).

757 The results were narrowed to 1999-2018 and the initial export was 3,042 records that after
758 cleaning resulted to 2,823 records. The content of the records was exported as Full Record in
759 the Tab Delimited UTF8 WIN file format.