

No Deal: German Researchers’ Publishing and Citing Behaviours after Failed Big Deal Negotiations with Elsevier

Abstract

In 2014, a union of German research organisations established Projekt DEAL, a national-level project to negotiate licensing agreements with large scientific publishers. Negotiations between DEAL and Elsevier began in 2016, and broke down without a successful agreement in 2018; in this time, around 200 German research institutions cancelled their license agreements with Elsevier, leading Elsevier to restrict journal access at those institutions from July 2018 onwards. We investigated the effect of these Big Deal negotiations and access restrictions on researchers’ publishing and citing behaviours from a bibliometric perspective, using a dataset of ~400,000 articles published by researchers at the affected DEAL institutions between 2012-2020. We further investigated these effects with respect to the timing of contract cancellations with Elsevier, research disciplines, collaboration patterns, and article open-access status. We find evidence for a decrease in Elsevier's market share of articles from DEAL institutions, from a peak of 26.3% in 2015 to 20.8% in 2020, with the largest year-on-year market share decreases (in excess of 1% per year) occurring from 2018 to 2020 following the implementation of access restrictions. We also observe year-on-year decreases in the proportion of citations made from articles published by authors at DEAL institutions to articles in Elsevier journals post-2018, although the decrease is smaller (-0.3% in 2019 and -0.6% in 2020) than changes in publishing volume. We conclude that failed Big Deal Negotiations with Elsevier and access restrictions have led to some reduced willingness of researchers at DEAL institutions to publish their research in Elsevier journals, but that researchers are not strongly affected in their ability to cite Elsevier articles, with the implication that researchers use a variety of other methods (e.g. interlibrary loans, sharing between colleagues, or “shadow libraries”) to access scientific literature.

Declarations

Author Contributions Conceptualisation: NF; Data curation: NF, AH; Formal analysis: NF, AH; Funding acquisition: PM, IP, NJ; Investigation: NF, AH; Methodology: NF, AH; Project administration: PM, IP; Software: NF; Supervision: PM, IP, NJ; Validation: NF, AH; Visualisation: NF; Writing - original draft: NF; Writing - review and editing: NF, AH, NJ, PM, IP.

Competing interests The authors declare no competing interests. During the initial preparation and submission of this manuscript, NF was an employee of ZBW - Leibniz Information Centre for Economics.

Following submission, NF changed employer and is currently an employee of Digital Science. Manuscript revisions were led by AH, and Digital Science had no input into the study at any time.

Acknowledgments This work was supported by the German Federal Ministry of Education and Research within the funding stream “Quantitative research on the science sector”, projects OASE (grant numbers 01PU17005A and 01PU17005B) and OAUNI (grant numbers 01PU17023A and 01PU17023B). We are grateful to Dimensions (<https://www.dimensions.ai/>) for providing API access through their scientometric research access program.

Data and code availability Aggregated datasets presented in this manuscript, as well as all code used for the data extraction, analysis, and manuscript preparation are available on GitHub (https://github.com/nicholasfraser/Projekt_DEAL) and archived on Zenodo (<https://doi.org/10.5281/zenodo.4771576>).

Introduction

In 2014 the [Alliance of Science Organisations](#) (*Allianz der Wissenschaftsorganisationen*; AWO), a union of the majority of German research organisations, established a national-level project named “[Projekt DEAL](#)” (herein referred to as “DEAL”), to negotiate licensing agreements for access to electronic journals of large scientific publishers. The key objectives of DEAL are:

1. To receive permanent, full-text access to the entire journal portfolio of the selected publishers.
2. To make all articles published by authors at German institutions automatically Open Access (OA).
3. To secure reasonable pricing according to a simple, future-oriented model based on publication volumes.

To date, DEAL negotiations have centred on three major publishers: Elsevier, Springer Nature and Wiley. Between 2012 and 2020, these three publishers were collectively responsible for publishing ~53% of scientific articles with at least a single author from a German research institution (Elsevier ~21%, Springer Nature ~21%, Wiley ~12%; data according to [Dimensions](#)). Negotiations between DEAL and Elsevier officially began in 2016, with Springer Nature and Wiley negotiations beginning a year later in 2017. In January 2019, DEAL announced the signing of an agreement with Wiley, fulfilling the defined negotiating objectives by allowing full access to Wiley’s portfolio of journals for institutions represented by DEAL (herein referred to as “DEAL institutions”), and automatic OA publishing of articles from corresponding authors at DEAL institutions under Creative Commons (CC) licenses, for an annual fee equal to €2,750 per published article ([Sander et al., 2019](#)). In January 2020 a similar agreement between DEAL and Springer Nature was signed, including the same per-article fee equal to €2,750 ([Kieselbach, 2020](#)).

Whilst negotiations with Wiley and Springer Nature have now successfully concluded in agreements, negotiations with Elsevier remain unresolved. At the end of 2016, ~70 German institutions¹ chose not to renew their contracts with Elsevier, leading to Elsevier restricting access to new journal issues at those institutions (and also restricting access to back-catalogues at some institutions) from the beginning of 2017 ([Vogel, 2017a](#)), although access was restored six weeks later ([Vogel, 2017b](#)). At the end of 2017, a further ~110 German institutions² decided not to renew their contracts with Elsevier, and at the beginning of July 2018, the [German Rectors’ Conference](#) (*Hochschulrektorenkonferenz*; HRK), who are leading negotiations on behalf of AWO, [announced](#) the breakdown and cancellation of all ongoing negotiations with Elsevier. In July 2018, authors at institutions which had cancelled their contracts with Elsevier had their access to new journal issues completely cut-off ([Else, 2018](#)). A further ~25 institutions, including the Max Planck Society and Fraunhofer Society³ did not renew their contracts with Elsevier at the end of 2018, bringing the total number of institutions without a contract with Elsevier to more than 200.

¹Archived webpage: https://web.archive.org/web/20191212094238/https://www.projekt-deal.de/vertragskuendigungen_elsevier/

²Archived webpage: <https://web.archive.org/web/20191212094842/https://www.projekt-deal.de/vertragskundigungen-elsevier-2017/>

³Archived webpage: <https://web.archive.org/web/20191114074949/https://www.projekt-deal.de/vertragskundigungen-elsevier-2018/>

As a provider of a large proportion of research published and cited by researchers at German institutions, failed negotiations and restricted access to Elsevier's article collections may have measurable effects on their publication and citation behaviour. Attempts to quantify such effects have already been made through recent survey approaches: A survey commissioned by Elsevier⁴ found that 61% of researchers at German institutions agreed or strongly agreed that losing access made their research activities less efficient, whilst 54% agreed or strongly agreed that losing access delayed the speed that they produce their research output. A separate survey of 384 researchers at the Faculty of Medicine of the University of Münster showed an overall similar sentiment, with 66% of researchers reporting that they now require more time to acquire literature and 46% of researchers reporting that losing access was a competitive disadvantage, yet only 29% of researchers reported that they would no longer write or review articles for Elsevier journals⁵. Following the implementation of access restrictions, a number of researchers also resigned from their positions on editorial boards of Elsevier journals⁶ (Vogel, 2017c).

The situation in Germany is not unique and breakdowns in negotiations between library consortia and Elsevier have been reported elsewhere. In Sweden, a number of universities, research institutes and government agencies were cut off from Elsevier journals between mid-2018 and the end of 2019 due to a breakdown in negotiations between Elsevier and the [Bibsam Consortium](#) (the national-level license negotiating body for Sweden). An [agreement](#) was eventually signed between Bibsam and Elsevier at the end of 2019, to take effect from 1st January 2020. A survey of 4,221 Swedish researchers carried out by Bibsam during the time period that Elsevier journals were inaccessible found that 51% of respondents were negatively affected in their desire to publish with Elsevier, and 54% had their work negatively impacted (Olson et al., 2020). The University of California (UC) system also had access to Elsevier journals restricted following the suspension of negotiations in 2019. A poll of ~7,300 UC affiliates (including faculty, graduate students, undergraduates, postdocs and other staff) during this time period found that 33% of respondents reported a significant impact of the loss of access, most greatly from the health sciences (52% reported a significant impact), but only 15% reported that it would affect their decision to publish in Elsevier journals⁷. In March 2021, the University of California [announced](#) a 4-year agreement with Elsevier, which, for the first time, also covers OA publication in the Cell Press and Lancet family of journals.

In parallel to these library-led negotiations, other researcher-led protests against the business practices of Elsevier have occurred in recent years. [The Cost of Knowledge](#) is a campaign launched by mathematician Timothy Gowers in 2012, asking researchers to sign a statement that they would refrain from publishing, refereeing or doing editorial work for Elsevier. To date, more than 18,000 researchers have publicly signed the statement. However, an analysis of the signatories in 2016 found that 38% of those who committed to

⁴Archived webpage: https://web.archive.org/web/20210427160017/https://resources.mynewsdesk.com/image/upload/fl_attachment/sgmm6zhwdmam4ys7evvf

⁵Archived webpage: <https://web.archive.org/web/20210429122105/https://www.uni-muenster.de/ZBMed/aktuelles/27850>

⁶Archived webpage: <https://web.archive.org/web/20210427081142/https://www.projekt-deal.de/aktuelles-zu-elsevier/>

⁷Archived webpage: <https://web.archive.org/web/20210518094220/https://www.library.ucsb.edu/uc-and-elsevier>

not publish in Elsevier journals actually published papers in Elsevier journals following their commitment (Heyman et al., 2016). Other protests have been made in the form of editorial board resignations; in 2015 the entire editorial board of *Lingua*, an Elsevier journal, resigned and formed a new journal named *Glossa* with a different publisher; a similar situation unfolded in 2019 when the editorial board of the *Journal of Informetrics* resigned and formed a new journal named *Quantitative Science Studies*. However, in both cases, the existing Elsevier journals continued to operate and publish new journal issues with newly-established editorial boards.

In this study, we aim to investigate researchers’ publishing and citing behaviours after Big Deal negotiations with Elsevier and restricted access to Elsevier journals at DEAL institutions in Germany. Despite suggestions from surveys that researchers were negatively affected in their desire to publish in Elsevier journals and negatively impacted in their daily research activities, to our knowledge no empirical evidence suggesting that this has actually occurred has been established. Specifically, we aim to answer the following research questions:

1. Did Big Deal negotiations with Elsevier and restricted access to Elsevier journals at DEAL institutions result in a change in researchers’ publishing behaviour?
2. Did Big Deal negotiations with Elsevier and restricted access to Elsevier journals at DEAL institutions result in a change in researchers’ citing behaviour?

For both of these research questions, we also consider variability with respect to the timing of contract cancellations of individual institutions, research disciplines, collaboration patterns, and article OA status.

Methods

Data Sources

DEAL institutions We collected names and contract expiration dates of 210 German universities, research institutions, higher education institutions and regional libraries that had their access to Elsevier articles restricted as part of the DEAL negotiations, using publicly available lists of institutions that cancelled their contracts with Elsevier in 2016⁸, 2017⁹ and 2018¹⁰ available on the [DEAL website](#). We manually mapped institutions to identifiers in the [Global Research Identifier Database](#) (GRID), using the available search interface on the GRID website. Of the 210 institution names provided, 209 were matched to a GRID identifier, with the exception of “HS Villingen-Schwenningen”, for which we were unable to unambiguously determine the correct GRID identifier. The list extracted from the Project DEAL website contained 10 individual records for each campus associated with the Baden-Wuerttemberg Cooperative State University

⁸Archived webpage: https://web.archive.org/web/20191212094238/https://www.projekt-deal.de/vertragskündigungen_elsevier/

⁹Archived webpage: <https://web.archive.org/web/20191212094842/https://www.projekt-deal.de/vertragskündigungen-elsevier-2017/>

¹⁰Archived webpage: <https://web.archive.org/web/20191114074949/https://www.projekt-deal.de/vertragskündigungen-elsevier-2018/>

(“Duale Hochschule Baden-Württemberg / DHBW”), but all campuses are collectively associated with a single identifier in the GRID database (“grid.449295.7”). The Max Planck Society and Fraunhofer Society were listed individually on the DEAL website, but both are umbrella associations that consist of a number of individual research institutions. Thus, we also extracted the GRID information for all their constituents, according to “parent-child” relationship information stored in GRID. The Helmholtz Association and Leibniz Association are similar umbrella associations, but the lists from the DEAL website contained the names of individual Helmholtz Association and Leibniz Association institutions, thus we limited the dataset to those contained directly in the list and did not extract information for all other constituent institutions. A limitation of this approach is that we rely solely on publicly available information of contract negotiations and cancellations; nonetheless, we attempted to verify the information contained on the DEAL website by manually searching for press releases or other informational web pages issued or maintained by the individual institutions that referred to restricted access to Elsevier articles. Of the original 210 institutions on the list, we found relevant information for 121 institutions; the information we discovered showed strong concordance with the information contained on the DEAL website (e.g. in terms of contract status and timing of restricted access) and thus we are confident in the quality of the publicly-available information.

Article and author metadata Article and author metadata used in this study were derived from three main bibliometric data sources: [Dimensions](#), [Crossref](#) and [Unpaywall](#). Initially, we retrieved article DOIs, complete author and affiliation details, fields of research, and reference lists (DOI-DOI links) for all articles with at least a single author from a DEAL institution, via the Dimensions Analytics API. These data were retrieved in the first two weeks of April 2021. Articles were limited to those with a publication date in years 2012 to 2020, and to “article” publication types. As the Dimensions Analytics API only allows a maximum of 50,000 records to be returned in a single query, we queried iteratively through each year and individual DEAL institution, using the associated GRID identifier, and extracted details of all articles that included an author at the respective institution. In a final step we combined all article records together and removed duplicates (e.g. where an article had authors from multiple DEAL institutions). Following these steps we created a set of 892,169 unique articles (overview of distribution over time and by publisher in Figure 1A), representing ~2.5-3% of total global output over the same time period (Figure 1B).

Figure 1C shows the distribution of the number of authors per article in our original dataset. The figure shows that a high proportion of articles contain a large number of authors. In cases of articles written by large teams or consortia, the contribution of DEAL authors to the writing of the article or subsequent publication strategy may be small. Figure 1E shows the proportion of articles in our original dataset in each year further subdivided by authorship types: “DEAL First Author” refers to articles where the first author is from a DEAL institution but the last author is not, “DEAL Last Author” where the last author is from a DEAL institution but the first author is not, “DEAL First and Last Author” where both the first and last authors are from DEAL institutions, and “DEAL Middle Author” where neither the first nor last authors

are from DEAL institutions. These results show that the majority of articles in our dataset have a first or last author (or both) from DEAL institutions, yet there exist a number of articles where DEAL authors are only included as middle authors. Although practices for the assignment of author order are neither clear nor consistent across disciplines ([Brand et al., 2015](#)), for the purposes of this study we make the assumption that the publication strategy for the article is primarily determined by either the first or last author of an article. Thus, as our study aims to focus on the direct behaviour of researchers at DEAL institutions, we subsequently limited our dataset to articles with a first AND last author from a DEAL institution (i.e. the group “DEAL First and Last Author” in Figure 1E)¹¹.

For each article retrieved from Dimensions, we also retrieved and parsed a complete list of references (total number of DOI-DOI reference links: 33,652,274). An overview of the distribution of references per article for our original dataset is shown in Figure 1D. We observed that an anomalously high number of articles in this distribution contained either zero references or a single reference. A manual investigation on a small random sample of these articles revealed that articles with zero references often represent diverse types of editorial content (e.g. corrections, errata, tables of content) which is indexed in Dimensions as “article” types. Articles with a single reference often represent abstracts which are linked to a single journal article (see, for an example, abstracts published by the journal [ChemInform](#)). In a small number of cases, articles that did not contain references in our dataset did in fact contain a reference list on the journal page - highlighting a potential weakness in Dimensions as a data source. However, as a broad generalisation, we conclude that articles containing zero references or a single reference do not represent “true” research articles, which is the focus of this study¹². We thus decided to remove these articles as a source of uncertainty from our dataset for subsequent analyses. Unfortunately, we discovered a major inconsistency in the availability of references in articles published by the American Chemical Society (ACS). While multiple references are present for almost all articles published by ACS before 2018, the majority of items with publication year 2018 and later had no references associated with them¹³. We therefore decided to exclude all items associated with this publisher from the analysis.

Following the removal of articles without DEAL first and last authors, articles with zero references or a single

¹¹An alternative approach here would have been to use “corresponding author” information rather than author positions; however, an analysis of a subset of articles from 10 randomly selected institutions (N articles = 6,513) revealed that only 57% of articles contained any corresponding author information, and coverage may be biased towards certain publishers. Of articles in this subset, we found that in 99% of cases where the first and last author were from a DEAL institution, that the corresponding author was also from a DEAL institution; although we may remove more articles from the analysis in our approach, we do not introduce any potential biases that may result from using incomplete corresponding author information.

¹²Unpaywall also provide an indicator of potential editorial content (defined in their data schema as “is_paratext”) which is based on the prevalence of keywords in titles (e.g. “editorial board” or “front cover”). Our method identifies a larger number of “non-research” articles than Unpaywall (12,284 versus 961), but of the 961 articles identified by Unpaywall, 99.8% are also removed via our method.

¹³Among the 30 largest publishers in our original dataset ACS was the only one with a majority of items indexed as articles not having any references associated with them from 2018 onwards. The year that this occurs is crucial for the analysis, since it is the year of cut-off from access for many institutions. Manual inspection of a sample showed that of 100 items with DEAL first and LAST author published by ACS that had zero or no references in our dataset 42 were indeed editorial content, abstracts or similar paratexts. The remaining items were articles or reviews with reference lists on the webpage or paywalled content which seemed to be “true” articles. Most of the ACS-items with DEAL first and last author and without references in our dataset did have references registered with Crossref.

reference, and articles by ACS, our final dataset of articles from Dimensions was reduced to 397,420 articles (45% of the original dataset).

For determining subject classifications, we used the “Fields of Research” (FOR) scheme available in Dimensions, which is itself based on the [Australian and New Zealand Standard Research Classification \(ANZSRC\)](#) system. The classification scheme consists of 22 divisions at the upper level. Unlike classification systems in other bibliometric databases which classify articles on a journal level, Dimensions firstly classifies articles on a single document level using a text-based classification approach. Where information is insufficient, Dimensions falls back to a journal-level classification. Some initial discussion of the strengths and weaknesses of the Dimensions approach has been conducted by [Bornmann \(2018\)](#), and [Herzog and Kierkegaard Lunn \(2018\)](#). [Bornmann \(2018\)](#) noted a number of inaccuracies in the classification of his own publication record in Dimensions. However, improvements to the classification system have since been implemented and Dimensions [reported](#) an increase in the precision and recall of the method in August 2019.

Article records from Dimensions were matched to records in Crossref (for classification of Elsevier versus Non-Elsevier content, using the Crossref member ID of Elsevier, [78](#)) and Unpaywall (for determination of article OA status). Matching was conducted through exact matching of DOIs: 99.7% and 99.8% of articles in our dataset from Dimensions were subsequently matched to articles in Crossref and Unpaywall, respectively. Crossref data is based on an openly available Crossref database snapshot ([Crossref, 2021](#)) that contains all Crossref records registered until 7th January 2021. Relevant metadata fields were parsed applying the *rcrossref* parsers ([Chamberlain et al., 2020](#)), following the same approach documented in [Jahn et al. \(2021\)](#). To reduce computation time and storage demands, the Crossref dataset we subsequently used for matching was limited to records registered after 1st January 2008 (*issued_date*). Since we focus on articles published from 2012 onwards, this will not affect our overall results. Unpaywall data is based on an openly available database snapshot (details available [here](#)) from February 2021. Processing of the Unpaywall dataset followed the same procedure as that documented in [Hobert et al. \(2021\)](#).

Data Processing, Storage and Analysis To allow fast data processing and analysis, all large datasets described above (i.e. those from Dimensions, Crossref and Unpaywall) were imported to [Google BigQuery](#), a cloud data warehouse which allows querying of large datasets with SQL. All analysis of data was subsequently carried out in R ([R Core Team, 2020](#)), using the DBI ([R Special Interest Group on Databases et al., 2021](#)) and bigrquery ([Wickham and Bryan, 2020](#)) packages to interface Google BigQuery directly with R. Throughout this mostly automated data gathering and analysis process, tools from the Tidyverse ([Wickham et al., 2019](#)) collection of packages for R were used.

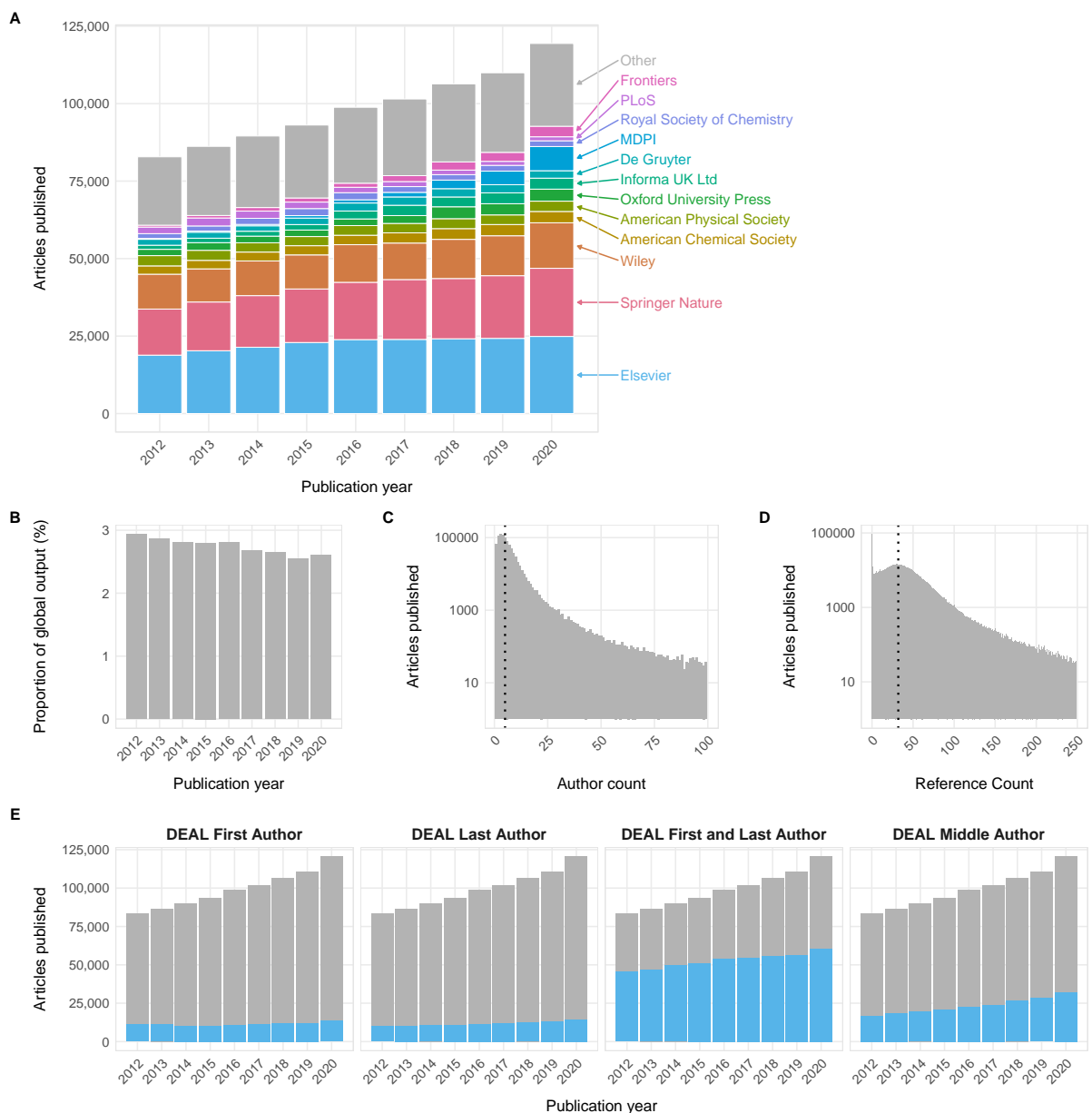


Figure 1: Properties of our initial dataset of articles published by DEAL researchers extracted from Dimensions. (A) Number of articles published per year and by publisher (top-12 publishers by total publication volume shown). (B) Proportion of articles in our dataset as a proportion of total global output per year (according to total number of articles indexed in Dimensions). (C) Distribution of number of authors per article (only articles with <100 authors are shown). Note y-axis is on a log-scale. The dotted line refers to median number of authors per article (5). (D) Distribution of number of references per article (only articles with <250 references are shown). Note y-axis is on a log-scale. The dotted line refers to median number of references per article (32). (E) Proportion of articles by authorship type: “DEAL First Author” refers to articles where the first author is from a DEAL institution but the last author is not, “DEAL Last Author” where the last author is from a DEAL institution but the first author is not, “DEAL First and Last Author” where both the first and last authors are from DEAL institutions, and “DEAL Middle Author” where neither the first nor last authors are from DEAL institutions.

Results

Publishing behaviour of DEAL researchers

In this section we assess possible changes in publishing patterns of researchers at DEAL institutions following Big Deal negotiations with Elsevier and restricted access to Elsevier journals. Whilst access restrictions have reduced the ability for these researchers to read and download Elsevier articles, there exist no further barriers for them to *publish* in Elsevier journals beyond those that previously existed, e.g. meeting submission and peer-review criteria, affordability of journal-specific fees, etc. However, we hypothesise that Big Deal negotiations and access restrictions may lead to negative sentiment amongst researchers which would influence their decision when choosing a suitable publication venue for their work; such negative desire to publish with Elsevier was reported by 51% of respondents of a survey conducted by the Bibsam Consortium when Elsevier restricted access to their journals in Sweden (Olson et al., 2020).

We assess changes in publishing behaviour of DEAL researchers primarily through two related metrics: (1) the total number of articles published in Elsevier versus non-Elsevier journals each year, and (2) the year-on-year (YOY) change in the absolute proportion of articles published in Elsevier journals (i.e. the change in Elsevier's market share of DEAL articles)¹⁴. With respect to both of these metrics, we consider variation with respect to the year of contract cancellation of individual DEAL institutions, research disciplines, collaboration patterns and article OA status. We focus on the period between 2012 and 2020, allowing us to capture the long-term trends in the years prior to Big Deal negotiations and access restrictions, and their effect on publishing patterns in the subsequent years.

Figure 2A shows the change in the number of articles published by DEAL researchers between 2012 and 2020 (limited to articles with DEAL first and last authors, with >1 reference, and publishers other than ACS). The total number of articles published per year increased during this period, from 37,156 articles in 2012 to 51,144 articles in 2020. In comparison, the number of articles published in Elsevier journals increased from 9,401 articles in 2012 to 11,651 articles in 2016, and subsequently decreased to 10,623 articles in 2020. In terms of Elsevier's market share of DEAL articles (Figure 2B), the years 2013-2015 show a trend of relatively small YOY market share gains (<0.6% per year), with Elsevier's market share reaching a peak of 26.3% in 2015. Subsequent years were characterised by a trend of larger market share losses, resulting in a final market share of 20.8% in 2020. These trends support the hypothesis that the desire of researchers at DEAL institutions to publish in Elsevier journals has been diminished following the start of Big Deal negotiations in 2016. The large drop in the number of articles published with Elsevier in 2018 (10,962 compared to 11,457 in 2017) and the corresponding YOY market share loss of -1.2% indicate that access restrictions in mid-2018 might have been a catalyst for this development. The biggest YOY market change appears in 2020 (-1.6%), however this is mainly the result of a large increase in the total number of publications which was not

¹⁴As an example, if DEAL researchers published 1000 articles in Year One, of which 200 were published in Elsevier journals, and 1500 articles in Year Two, of which 240 were published in Elsevier journals, then the change in proportion from Year One to Year Two is calculated as $(240/1500) - (200/1000)$, equal to -0.04 and interpreted as a year-on-year market share loss of 4%.

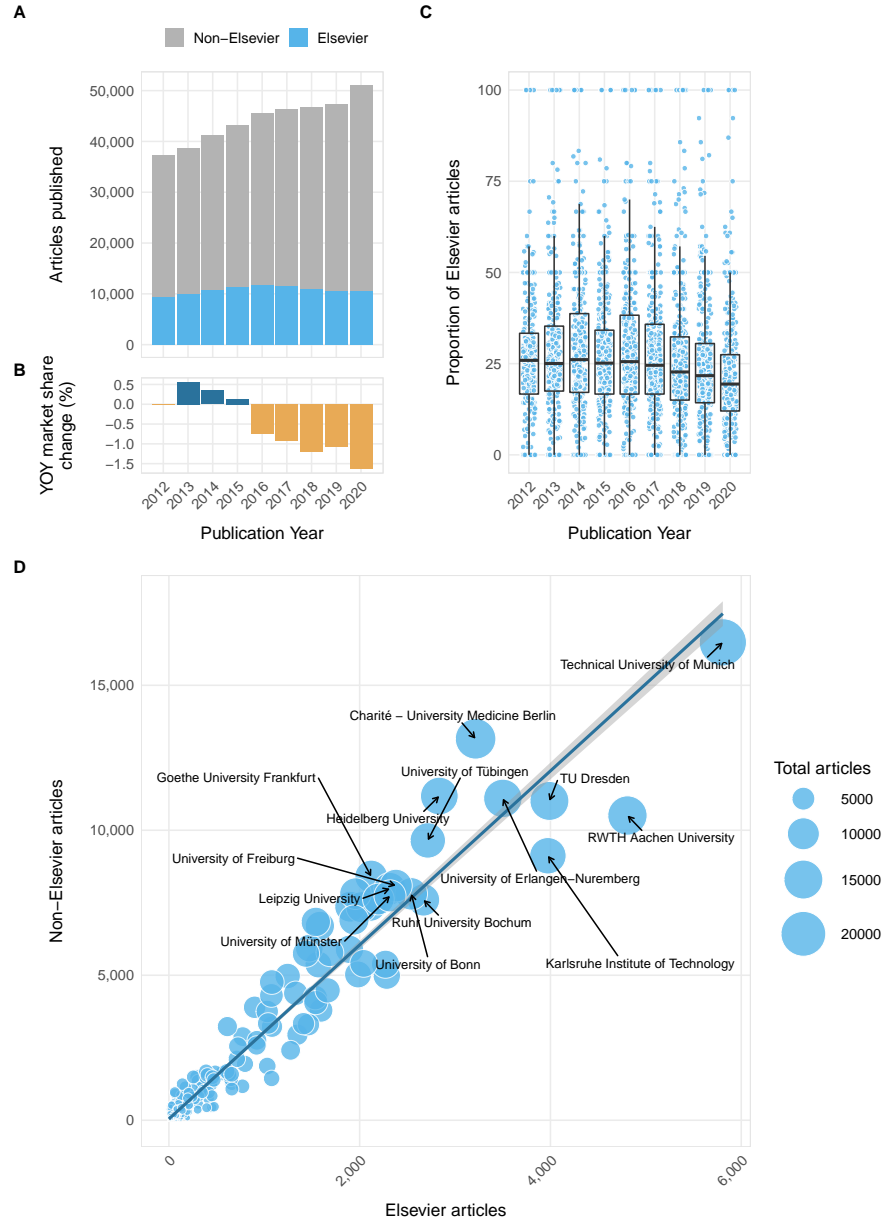


Figure 2: Publishing behaviour of DEAL researchers, 2012-2020. (A) Total number of articles published by DEAL researchers in Elsevier and non-Elsevier journals. (B) Year-on-year (YOY) change in Elsevier's market share of articles published by DEAL researchers. (C) Proportion of articles published by individual DEAL institutions in Elsevier journals. Boxplot horizontal lines denote lower quartile, median, upper quartile, with whiskers extending to $1.5 \times \text{IQR}$. Points denote individual institutions, with added horizontal jitter for visibility. (D) Number of Elsevier versus Non-Elsevier articles published for each individual DEAL institution (totals aggregated for years 2012-2020). Point size is scaled by the total number of articles published by an institution. Institutions that published $>10,000$ articles in total are labelled.

matched by publications in Elsevier journals. It is important to note that these results reflect the numbers and proportions of articles *published* in journals, but do not necessarily reflect article *submission* dynamics: articles take many months to proceed through peer-review and publication processes (and these processes are generally faster in STM fields versus social sciences/arts/humanities/economics fields; Björk and Solomon, 2013), and acceptance/rejection rates may not have remained static or proportional over time.

Figure 2C shows changes in the proportion of articles from each individual DEAL institution that are published in Elsevier journals. Patterns broadly reflect those shown in Figures 2A and 2B, with the proportion of articles in Elsevier journals remaining relatively static from 2012-2016, and subsequently declining. Interestingly, a small number of DEAL institutions appear to publish 100% of their articles in Elsevier journals - upon inspection we find that these reflect institutions with extremely low publication volumes (<10 articles in a given year). A similar picture can be seen on the other side of the spectrum: DEAL institutions with 0% of articles in Elsevier journals tend to have low total numbers of publications in that year (<5 articles). However, there are some exceptions of DEAL institutions with up to 59 publications in a year of which none are published in Elsevier journals. The majority of these institutions are Max Planck Institutes, or other non-university research institutions with a focus in Physics or other Natural and Computer Sciences. Figure 2D shows the total number of Elsevier versus non-Elsevier articles published by DEAL institutions aggregated over the entire time period 2012-2020. Results show general consistency in the proportion of articles published in Elsevier journals between large and small institutions; however, institutions of similar sizes also show sizeable variation. For example, Charité - University Medicine Berlin and RWTH Aachen University both published on the order of ~15,000 articles between 2012 and 2020, yet only ~20% of articles published by Charité were published in Elsevier journals compared to ~31% by RWTH. Such large differences may reflect different research focuses of individual institutions, e.g. Charité has a strong biomedical focus, whilst RWTH is a technical university with a historically strong focus in natural sciences, technology and engineering.

Did publishing patterns differ at institutions whose contracts with Elsevier expired in different years? Our dataset contains three groups of DEAL institutions whose contracts with Elsevier expired at different time points: one group at the end of 2016 ($N = 74$), one at the end of 2017 ($N = 110$) and one at the end of 2018 ($N = 204$). In Figure 3 we explore differences in publishing patterns between these different groups, to examine whether those whose contracts expired earlier showed different effects to those whose contracts expired more recently. All three groups grew in total publication volume between 2012 and 2020; the largest group by publication volume was those whose contracts expired in 2017, and the smallest those whose contracts expired in 2018. With respect to the share of articles published in Elsevier journals, all three groups display similar dynamics with an overall gain in Elsevier's market share between 2012 and 2015, and an overall loss between 2016 and 2020, although the exact magnitude and patterns of YOY gains and losses differs between each group. Between 2015 and 2017 (i.e. two years prior to the access restrictions in

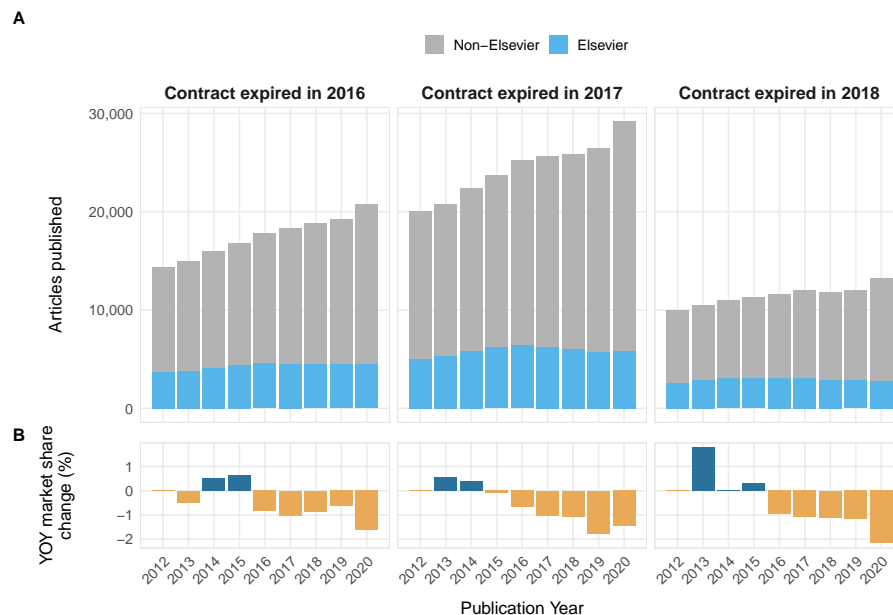


Figure 3: Changes in publishing behaviour of DEAL researchers, 2012-2020, dependent on year of contract expiration with Elsevier. (A) Total number of articles published by DEAL researchers in Elsevier and non-Elsevier journals. (B) year-on-year (YOY) change in Elsevier's market share of articles published by DEAL researchers.

2018), Elsevier's market share changed by -1.9%, -1.7% and -2% for the group whose contracts expired in 2016, 2017 and 2018, respectively; the rate of market share losses increased for all three groups between 2018 and 2020 (i.e. two years following the access restrictions in 2018) to -2.2%, -3.2% and -3.3%. Again, this points to a declining trend which is accelerated by the access restrictions to Elsevier journals. A reason for the relatively homogenous behaviour across all three groups may be that although the contracts expired at different timepoints, the time at which access to Elsevier was restricted was relatively similar across all DEAL institutions; those whose contracts expired at the end of 2016 and 2017 lost access in July 2018 (not including a brief 6-week period at the beginning of 2017), whilst those whose contracts expired at the end of 2018 lost access from the beginning of 2019 onwards.

How did publishing behaviour vary with respect to research disciplines? A complicating factor in our dataset is that we have included articles covering multiple research disciplines; we have already shown in Figure 2D that variation in publishing patterns exists on the institutional level, which may be a consequence of the individual institutions' research focuses. A recent analysis of the effect of the agreements made between DEAL and Springer Nature, and DEAL and Wiley (Haucap et al., 2021), chose to focus on a single discipline, Chemistry, with the justification that:

“Manuscript turnaround times differ substantially between different fields of science and are rather long in some disciplines such as economics (see, e.g., Ellison, 2002). Hence, the vast majority of articles published in economics journals in 2019 and 2020 will have been submitted before the

DEAL agreements were announced. Therefore, our analysis focuses on the field of chemistry which has much faster turnaround times so that we can expect the DEAL agreements to already have at least some impact.”

Our analysis covers a longer time period than that of [Haucap et al. \(2021\)](#), and we assess effects covering the entire time period in which negotiations with Elsevier began in 2016, the time at which access was restricted in July 2018, and two subsequent publication years thereafter until the end of 2020. We therefore feel justified in including a broader range of disciplines in our approach with longer publishing timelines than Chemistry. Nonetheless, we have also analysed changes at the level of individual disciplines (i.e. Dimensions Fields of Research). A visualisation of yearly developments of publication volumes, Elsevier's market shares and YOY market share changes for all 22 disciplines can be found in the (updated) supplementary material archived on Zenodo: <https://doi.org/10.5281/zenodo.4771576>. Since the discipline-specific analysis did not show clear patterns, we do not include the corresponding figures here.

Overall, patterns of publishing behaviour for individual research disciplines display a higher degree of fluctuation than our overall analysis with all disciplines aggregated (likely due to smaller sample sizes). Most disciplines show overall growth in the total number of articles published over the timeframe of our analysis, although publication volumes in some disciplines appear to have decreased or reached a plateau in recent years (e.g. Chemical Sciences, Physical Sciences). The results also reveal variation in the tendency of authors to publish in Elsevier journals between disciplines: aggregated over the entire period between 2012 and 2020, the discipline with the highest proportion of articles published in Elsevier journals is Economics (41.9%), and the lowest proportion is in Philosophy and Religious Studies (5.5%). In 19 of the 22 disciplines, Elsevier lost overall market share between 2012 and 2020; the largest losses are reported in Earth Sciences (−14.6%), Language, Communication and Culture (−13%), Chemical Sciences (−12.4%), Agricultural and Veterinary Sciences (−11.5%), and History and Archaeology (−11.4%); conversely, three individual disciplines showed market share gains over the same time period (Economics, +2.8%; Commerce, Management, Tourism and Services, +4.5%; Technology, +5.8%). However, the patterns of YOY growth/losses in Elsevier's market share for individual disciplines vary significantly and we do not observe any consistent trends that can be clearly attributed to Big Deal negotiations or the access restrictions in 2018; only 10 of the 22 disciplines showed greater market share losses following Big Deal negotiations or the access restrictions. We point out that the results obtained for Chemical Sciences have to be interpreted with caution, since we excluded a major publisher (ACS) for this discipline in our analysis (see previous “Methods” section for details).

How did publishing behaviour vary with respect to collaboration patterns? Another potential confounder of our overall results in earlier sections relates to that of collaboration behaviour. An article written solely by researchers at DEAL institutions may have a different publication strategy compared to one written in an internationally-collaborative project, where international colleagues may be less knowledgeable of Elsevier Big Deal negotiations and access restrictions, or less disturbed in their daily research activities.

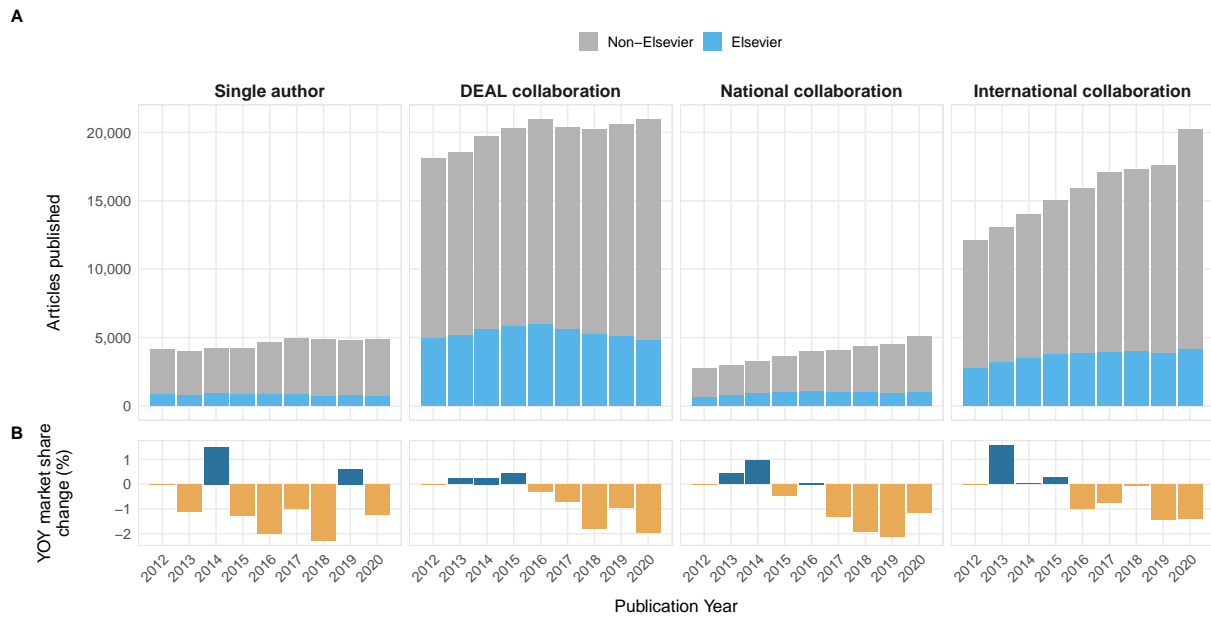


Figure 4: Changes in publishing behaviour of DEAL researchers, 2012-2020, dependent on collaboration status. “Single author” refers to articles that are authored by a single researcher, “DEAL collaboration” refers to articles with multiple authors, where all authors are based exclusively at DEAL institutions, “National collaboration” refers to articles where some authors are based at DEAL institutions and others at non-DEAL institutions within Germany, and “International collaboration” refers to articles where some authors are based at DEAL institutions and others at institutions outside of Germany. (A) Total number of articles published by DEAL researchers in Elsevier and non-Elsevier journals. (B) year-on-year (YOY) change in Elsevier's market share of articles published by DEAL researchers.

Figure 4 shows changes in publishing behaviour with respect to collaboration status. We classified articles into four distinct collaboration classes: (1) “Single author”, referring to articles published by only a single researcher, (2) “DEAL collaboration”, referring to articles with multiple authors where all authors of the article are based exclusively at DEAL institutions, (3) “National collaboration”, referring to articles where some authors are based at DEAL institutions and others at non-DEAL institutions within Germany, and (4) “International collaboration”, referring to articles where some authors are based at DEAL institutions and others at institutions outside of Germany. Note that for all of these classes, the first and last authors (or the single author) always have an affiliation at a DEAL institution. For all of the collaborative classes (i.e. ignoring the single-author class) the minimum number of authors per article was 2, the median number of authors was 3, 6 and 6 for DEAL collaborations, national collaborations and international collaborations, respectively, and the maximum number of authors was 68, 48 and 729 for DEAL collaborations, national collaborations and international collaborations, respectively.

Overall, most publications in our datasets are produced in DEAL collaborations or international collaborations, whilst the number of articles published by single-authors and in other national collaborations is comparatively small. The number of single-author articles has remained relatively static over time, only increasing from 4,135 in 2012 to 4,892 in 2020. With respect to DEAL-only collaborations, the total number of published articles grew from 18,121 in 2012 to 20,962 in 2016, but subsequently plateaued or even slightly declined, with 20,935 articles published in 2020. In comparison, the number of articles published as national collaborations and international collaborations grew substantially from 2012 to 2020 (from 2,761 to 5,060 for national collaborations, and 12,139 to 20,257 for international collaborations). The findings suggest that over time, DEAL researchers are transitioning towards a more collaborative research environment, particularly with respect to increasing collaborations with international partners, which echoes similar trends observed across Europe ([Kwiek, 2020](#)).

In terms of Elsevier's market share, single-author articles show a general long-term trend towards a market share loss over time (from an overall market share of 20.9% in 2012 to 14.2% in 2020), albeit punctuated by two YOY market share gains in 2014 and 2019. The largest YOY losses appear in 2016 (-2%) and 2018 (-2.3%), the years of start of negotiations and access restrictions, respectively. All three collaborative groups also display overall losses between 2012 and 2020 (from an overall market share of 27.8%, 25.5%, and 23.1% in 2012, to 23%, 20%, and 20.3% in 2020 for DEAL collaborations, national collaborations and international collaborations, respectively), yet the patterns and timing differ somewhat between different collaboration groups. The share of articles published in Elsevier journals as DEAL collaborations decreased relatively steadily from 2016 to 2020, with the steepest decreases in 2018 (YOY market share loss of -1.8%) and in 2020 (-1.9%). In comparison, the market share of articles published as national collaborations shows only slight fluctuations until 2016 and began to decrease in 2017, and the overall market share decrease for articles published as international collaborations between 2016 and 2020 was punctuated by a very small YOY

decrease of only -0.1% in 2018.

How did publishing behaviour vary with respect to OA status? Between 2010 and 2018 the proportion of articles authored by all researchers at German institutions that were made OA increased dramatically, from 27% in 2010 to 52% in 2018 (Hobert et al., 2021). We also aimed to determine whether the restriction of access to Elsevier journals had an effect on OA publishing behaviour of DEAL researchers in Elsevier journals, where we speculate that increased awareness of access issues, and motivation to ensure accessibility for colleagues, would motivate DEAL researchers to publish articles under OA publishing models. Articles were classified into OA categories following the same schema used by Unpaywall: broadly, “Gold” refers to articles published in fully OA journals, “Hybrid” to articles published under a OA license in an otherwise subscription-based journal, “Green” to articles that have been made available in an OA repository but access to the publication in the journal is restricted, and “Bronze” to articles that are freely accessible on the publisher’s website but are not published under an OA license. All articles that are not freely accessible are classified as “Closed”. An important point for the analysis of OA shares is that our dataset measures OA availability at the time of measurement (in our case, February 2021), and so OA shares do not necessarily reflect the OA status of an article at the time of its publication (e.g. an article could transition from Closed to Green several years after publication, if a version is deposited to an OA repository after an embargo period).

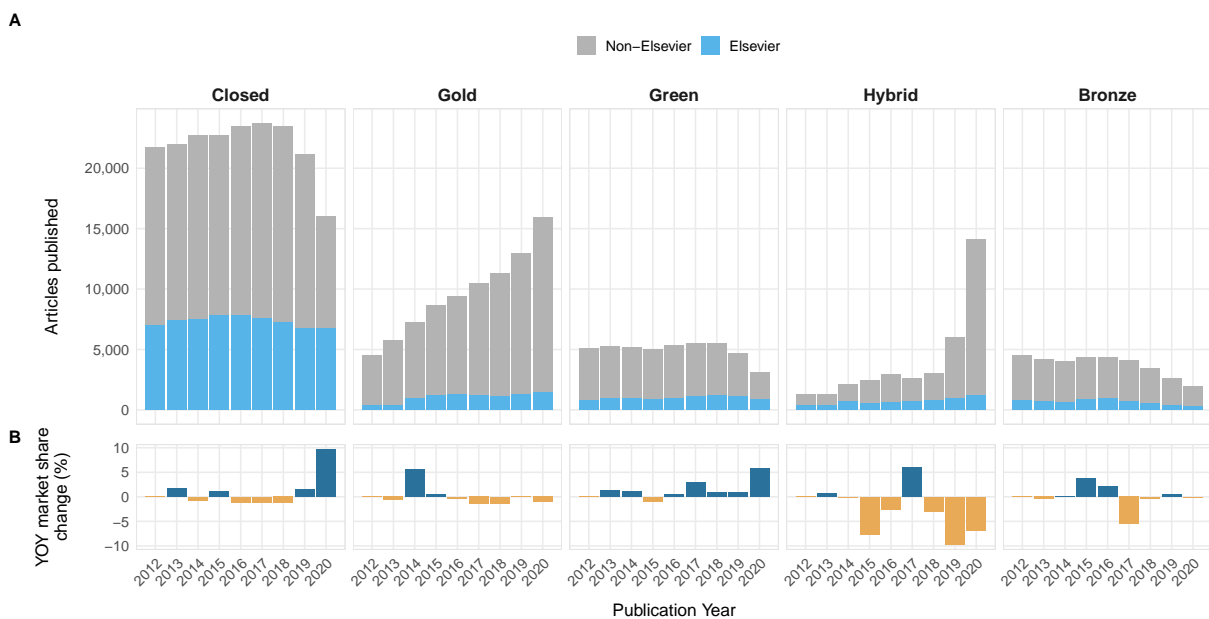


Figure 5: Changes in publishing behaviour of DEAL researchers, 2012-2020, dependent on OA status. (A) Total number of articles published by DEAL researchers in Elsevier and non-Elsevier journals. (B) year-on-year (YOY) change in Elsevier’s market share of articles published by DEAL researchers.

OA publishing patterns of DEAL institutions shown in Figure 5 broadly agree with the previous findings of Hobert et al. (2021) which focused on a larger number of German universities and non-university research

institutions. We find that the total proportion of OA articles (including all OA categories analysed) published by DEAL researchers grew from 41.5% in 2012 to 68.7% in 2020, compared to a growth in OA articles of 27% in 2010 to 52% in 2018 reported by [Hobert et al. \(2021\)](#). The strong growth of OA at DEAL institutions is driven largely by the growth of Gold OA from 2012 (4,544 articles) to 2020 (15,940 articles), and of Hybrid OA in particular from 2018 (3,016 articles) to 2020 (14,163 articles). The strong increase in Gold OA can be attributed mainly to the growth of MDPI and Frontiers as well as a larger number of smaller OA publishers in recent years, whereas the increase in Hybrid OA is driven almost exclusively by climbing numbers of publications with Springer Nature and Wiley¹⁵ where agreements were obtained (see also [Haucap et al., 2021](#)).

With respect to publishing patterns in Elsevier journals, the growth of OA has been relatively moderate compared to the overall picture in Germany: from 2012 to 2020 the total proportion of DEAL articles published in Elsevier journals that were made OA increased only from 25.2% to 36.8%. Regarding individual OA categories, we observe the largest changes occurring from around 2014 onwards in Elsevier journals: the number of Gold articles increased from 950 articles in 2014 to 1,497 articles in 2020, and the number of Hybrid articles increased from 681 articles in 2014 to 1,188 articles in 2020. In terms of market share, the most prominent feature is that of YOY losses for Elsevier in the Hybrid OA market share of >5% in consecutive years in 2019 and 2020, however, this appears to be driven more by the surge of Hybrid OA publishing in other venues rather than a reduction in the volume of Hybrid OA published by Elsevier (which is also reflected in the large gain of market share (+9.7%) for Elsevier of closed articles in 2020). Another interesting feature is the growth of Green OA in Elsevier journals in 2020 (+5.8% market share), suggesting an increasing proportion of researchers publishing in Elsevier journals are depositing their work to OA repositories in comparison to across all publishers as a whole. Bronze OA will not be discussed separately here, since the decision for Bronze OA is made on the side of the publisher, not the author, and hence it is not influenced by changes in author’s publishing choices.

Citing behaviour of DEAL researchers

A number of studies have found a citation advantage of open-access publications over their closed-access counterparts (c.f. [Piwowar et al., 2018](#)), with the implication that having access to articles makes them easier to read, download and ultimately more likely to be cited. If this were true, we would expect that restricting access to a set of articles would have the opposite effect, i.e. reduce their ability to be cited. We therefore aimed to investigate the effect of Elsevier access restrictions on DEAL researchers citing behaviour, using a set of 16,298,286 references (DOI-DOI links) from articles authored by DEAL researchers (limited to articles where the first and last author was at a DEAL institution, articles that contained more than a single reference, and articles not from ACS: see previous “Methods” section for details).

¹⁵Figures displaying number of articles published per year and OA category and by publisher (top-10 publishers by total publication volume in the respective OA category shown) can be found in the supplementary material on [Zenodo](#)

A complicating factor in the analysis of citing behaviour is that there exists time variation in both the year of publication of an article, and in the publication year of the articles they cite, i.e. articles may cite other articles published in any prior year. An overview of citing dynamics for our dataset of DEAL articles is displayed in Figure 6, where we show the mean number of citations per article made to Elsevier or non-Elsevier articles, as a function of citing year (where citing year refers to the difference in years between the publication date of the citing article and the publication date of the cited article). An important point for this section is that we study the problem in an inverse way to the majority of citation studies: we analyse *outgoing* rather than *incoming* citations. Nonetheless our results display similarly typical dynamics (see, e.g. [Parolo et al. \(2015\)](#)): articles in our dataset cite relatively few articles published in the same year (presumably, as citing articles must first be written and proceed through the lengthy peer-review process), the number of cited articles peaks in citing years 2-3, and subsequently slowly declines over the following years.

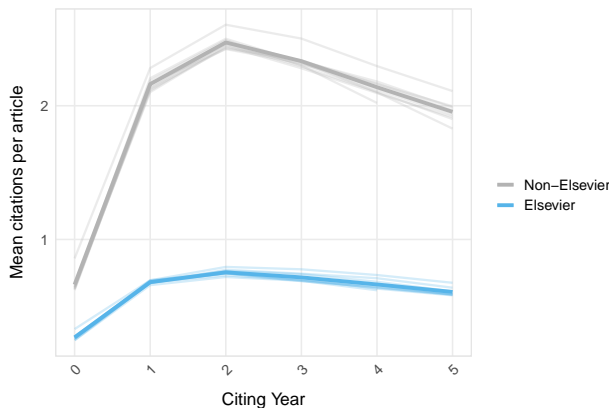


Figure 6: Citing dynamics for articles published by DEAL researchers. Lines show the evolution of the mean number of citations made to Elsevier or non-Elsevier articles, as a function of citing year. Thick lines represent the overall mean, thin lines represent articles from individual publication years.

For the purposes of the following analyses, we make the assumption that citing behaviour in response to Elsevier access restrictions is most likely to change for recent citations, which we define as citations with a citation age less than or equal to 2 years. This assumption is based on the fact that access restrictions affected new journal issues at all DEAL institutions, whilst only a subset of DEAL institutions also lost access to their back-catalogue of articles ([Vogel, 2017a](#)). We may therefore reasonably expect that authors who have had access restricted to Elsevier journals post-2018 either still have access to the older back-catalogues, or may have saved older articles to local storage media (e.g. in reference management software) during the time when access was still available. Hence, we restrict the citation analyses to citations with a citation age less or equal to 2. For assessing changes in citing behaviour, we define three related metrics of measurement in similar way to our analysis of publishing behaviour: (1) the proportion of articles published by DEAL researchers that cited *any* article in an Elsevier journal, (2) the total number of citations made to articles in Elsevier versus non-Elsevier journals from articles published by DEAL researchers, and (3) the annual change in the absolute proportion of citations made to articles in Elsevier journals from articles published by DEAL

researchers (which, for the purposes of this analysis, we will term as Elsevier's "market share" of citations).

Overall results of citing behaviour of DEAL researchers' between 2012 and 2020 are shown in Figure 7, as well as citing behaviour for individual institutions. All results are based on a maximum citation age of two years. Figure 7A shows the proportion of articles published in a given year, that cite at least a single Elsevier article. Overall, the proportion of articles that have cited an Elsevier article has remained relatively constant over time, fluctuating around a mean of 57.4%. Figure 7B shows that the total number of citations largely echoes trends of the number of articles published in the same year (Figure 2A); however, whilst the total number of Elsevier articles published by DEAL researchers decreased from 2015 to 2020 (Figure 2A), the number of citations made to Elsevier articles continued to rise, from 77,542 citations in 2016 to 93,035 citations in 2020, although this is largely driven by the large single-year peak in 2020. The result of these patterns is that over the entire time period of our analysis, Elsevier's market share of citations from DEAL researchers increased slightly from 23.4% in 2012 to 24% in 2020, with a prominent gain in market share in 2018 (+0.6%) and the largest market share loss occurring in 2020 (-0.6%). The losses of market share in 2019 and 2020 are consistent with an expectation that reduced access to Elsevier articles post-2018 would have reduced the ability of researchers to cite those articles; however, in comparison to market share losses in publishing volume over the same time period (-2.7%), the loss of market share of citations (-0.9%) is relatively moderate.

Figures 7D and 7E display citing behaviour at the level of individual institutions. Overall, the proportion of citations made to articles in Elsevier journals reflect the same patterns of publishing behaviour as observed in Figures 2C and 2D - the proportion of citations made to articles in Elsevier journals gradually increased from 2012 to 2018, and small decreases were noted in 2019 and 2020. Aggregated over the entire 2012-2020 time period, we find little evidence of size-related effects: large research institutions generally cite articles in Elsevier journals in similar proportions to small research institutions (Figure 2E), although some variation between individual institutions exists.

Did citing patterns differ at institutions whose contracts with Elsevier expired in different years? As with our analysis of publishing patterns (Figure 3), we also analysed citing behaviour of DEAL researchers dependent on the contract expiration date of their institution with Elsevier (Figure 8). Results show relatively homogenous behaviour between the different groups and reflect the overall results from Figure 7. One prominent feature, however, is a YOY market share loss of -1.5% in 2020 for the group whose contracts with Elsevier expired in 2018; in comparison, the groups whose contracts expired in 2016 and 2017 show only relatively small losses for the same year (-0.3% and -0.7%, respectively).

How did citing behaviour vary with respect to research disciplines? We analysed how citing behaviour also varied with respect to individual research disciplines (Dimensions Fields of Research). Full results for all 22 research disciplines are available in the supplementary material archived on Zenodo: <https://doi.org/10.5281/zenodo.4771576>.

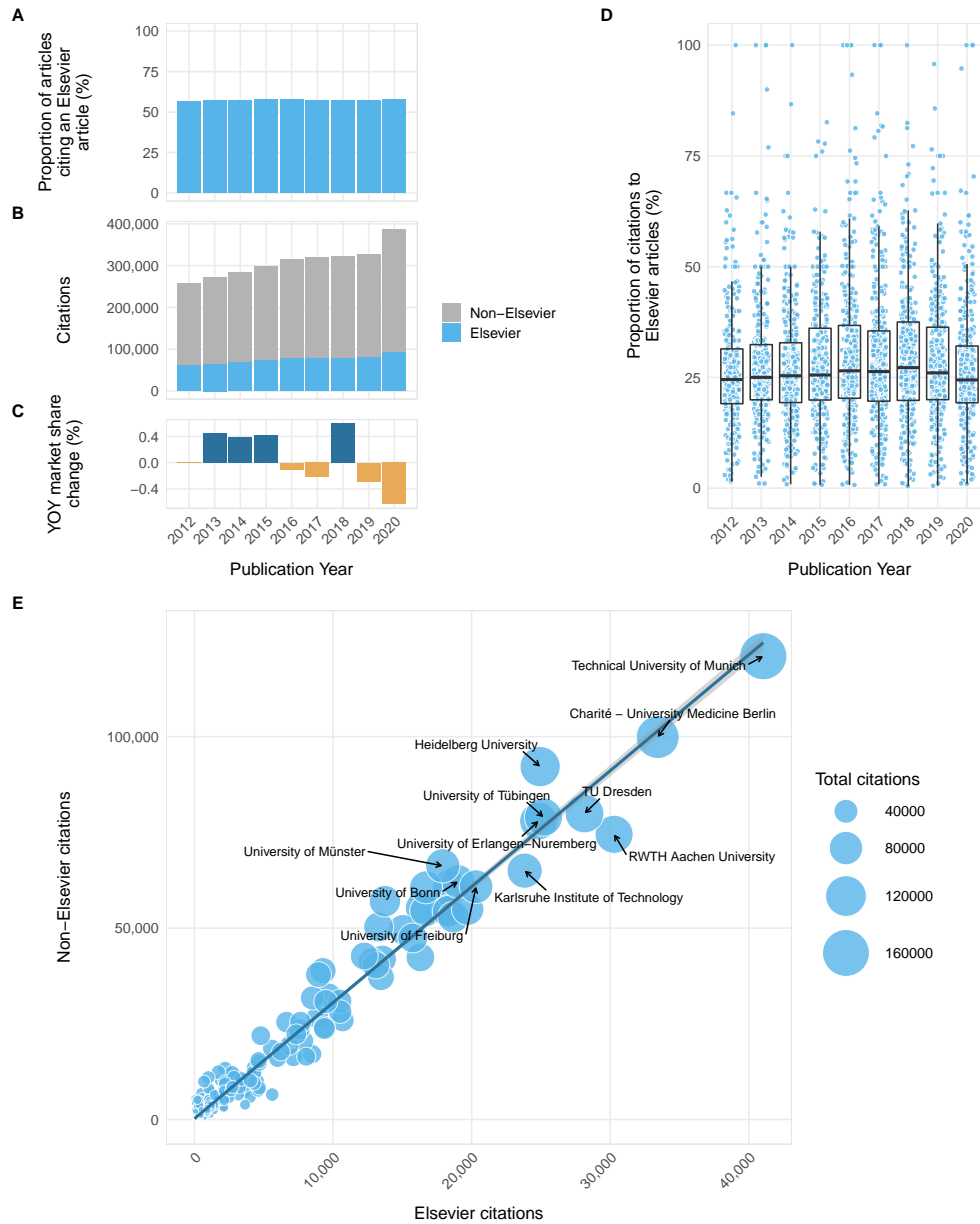


Figure 7: Citing behaviour of DEAL researchers, 2012-2020. All data is based on citations from articles published by DEAL researchers to articles with a citation age of 2 years or less. (A) Proportion of articles published by DEAL researchers that cite at least one single Elsevier article. (B) Total number of citations from articles published by DEAL researchers to articles published in Elsevier versus non-Elsevier journals. (C) Year-on-year (YOY) change in Elsevier’s market share of citations from articles published by DEAL researchers. (D) Proportion of citations from articles published by individual DEAL institutions to articles in Elsevier journals. Boxplot horizontal lines denote lower quartile, median, upper quartile, with whiskers extending to 1.5*IQR. Points denote individual institutions, with added horizontal jitter for visibility. (E) Number of citations to articles in Elsevier versus Non-Elsevier journals for articles published by individual DEAL institutions (totals aggregated for years 2012-2020). Point size is scaled by the total number of citations from an institution. Institutions citing $>80,000$ articles are labelled.

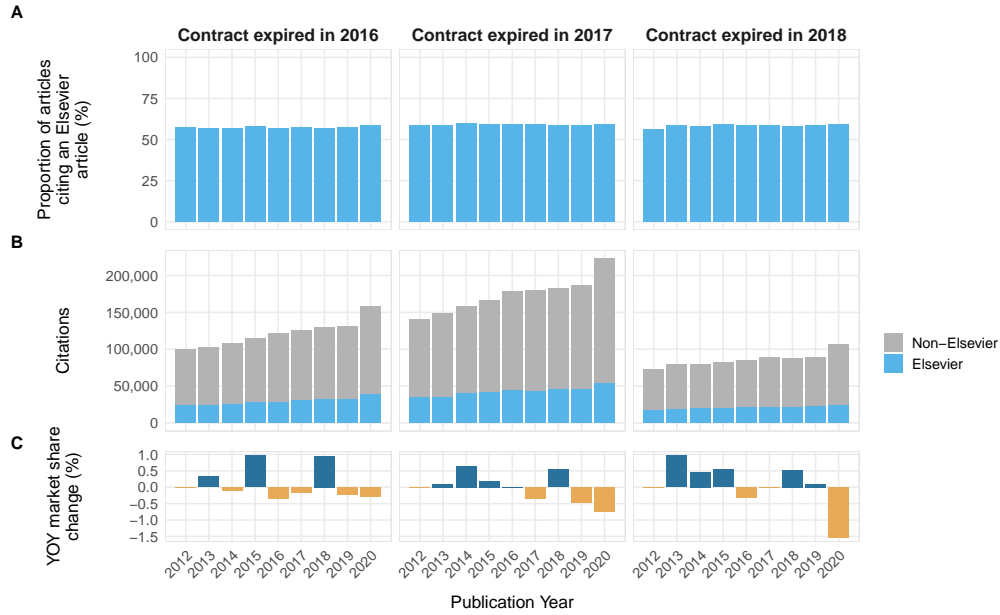


Figure 8: Citing behaviour of DEAL researchers, 2012-2020, dependent on year of contract expiration with Elsevier. All data is based on citations from articles published by DEAL researchers to articles with a citation age of 2 years or less. (A) Proportion of articles published by DEAL researchers that cite at least one single Elsevier article. (B) Total number of citations from articles published by DEAL researchers to articles published in Elsevier versus non-Elsevier journals. (C) Year-on-year (YOY) change in Elsevier's market share of citations from articles published by DEAL researchers.

We observe a high degree of variation between disciplines in their propensity to cite articles in Elsevier journals: aggregated across all publication years, the research disciplines with the highest proportion of citations to articles in Elsevier journals were Built Environment and Design (47.7%) and Economics (43%), whilst the disciplines that made the lowest proportion of citations to articles in Elsevier journals were Physical Sciences (12.2%) and Philosophy and Religious Studies (11.6%). We also analysed the proportion of articles that cited at least one Elsevier article for each individual discipline, and found the highest proportion in Environmental Sciences (74.1%, aggregated across all publication years), and lowest proportion in Philosophy and Religious Studies (15.1%). However, these results need to be carefully interpreted as the number of references per article (i.e. the reference density) also varies between research disciplines (Sánchez-Gil et al., 2018). Articles in disciplines where the average number of references per article is higher are likelier to cite at least a single Elsevier article, all other factors being equal.

Given the high variability in YOY market shares for individual disciplines, it is difficult to interpret clear long-term trends that can be attributed directly to the Elsevier access restrictions in 2018. However, we do find that from 2018 to 2020, the share of citations made to articles in Elsevier journals decreased in 15 of the 22 disciplines (with the largest loss of -10.2% in Built Environment and Design) - in only 7 disciplines did the share of references to articles in Elsevier journals increase during this time (Philosophy and Religious Studies, +1.5%; Studies in Creative Arts and Writing, +0.8%; Agricultural and Veterinary Sciences, +0.8%; Studies

in Human Society, +0.4%; Commerce, Management, Tourism and Services, +0.2%; Technology, +0.1%; Chemical Sciences, +0.1%).

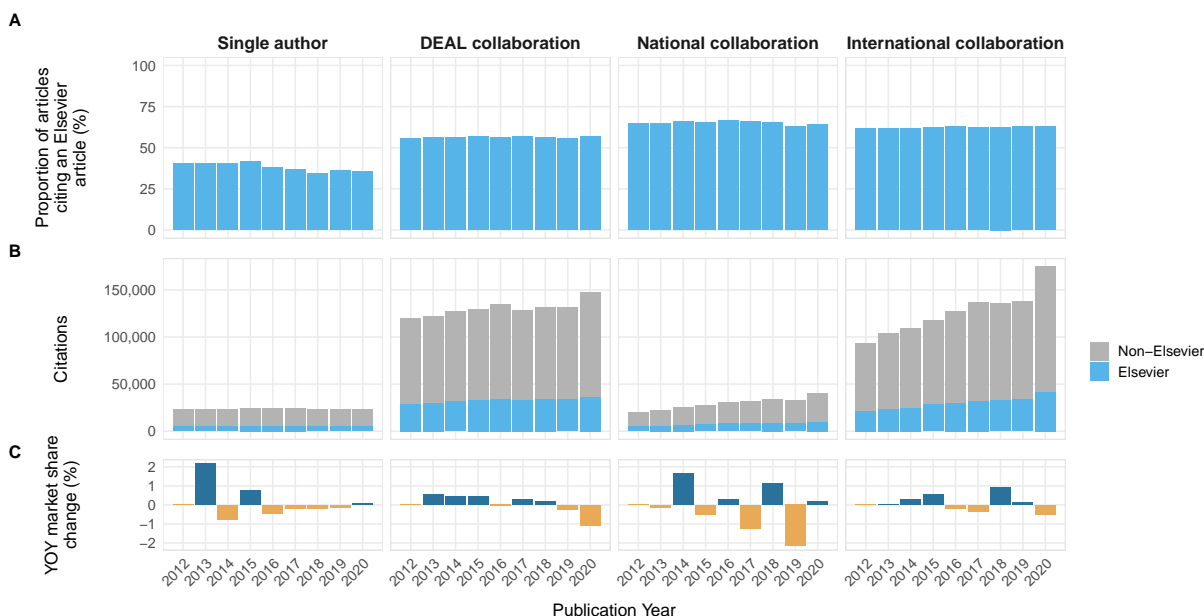


Figure 9: Citing behaviour of DEAL researchers, 2012-2020, dependent on collaboration status. “Single author” refers to articles that are authored by a single researcher, “DEAL” collaboration’ refers to articles with multiple authors, where all authors are based exclusively at DEAL institutions, “National collaboration” refers to articles where some authors are based at DEAL institutions and others at non-DEAL institutions within Germany, and “International collaboration” refers to articles where some authors are based at DEAL institutions and others at institutions outside of Germany. All data is based on citations from articles published by DEAL researchers to articles with a citation age of 2 years or less. (A) Proportion of articles published by DEAL researchers that cite at least one single Elsevier article. (B) Total number of references in articles published DEAL researchers to articles published in Elsevier versus non-Elsevier journals. (C) Year-on-year (YOY) change in Elsevier's market share of references in articles published by DEAL researchers.

How did citing behaviour vary with respect to collaboration patterns? Collaboration patterns may also have an influence on citing behaviour. Some evidence exists (e.g. [Milojević, 2012](#)) that authors who collaborate more extensively use more references, and tend towards “younger” references, compared to authors who collaborate less. Here, we may hypothesise that more collaborative articles (i.e. those not solely involving researchers at DEAL institutions) may receive more input into the writing and referencing process from researchers that are not subjected to the DEAL access restrictions, and thus any measurable effect on citing behaviour should be weakened in the more collaborative groups. Our analysis, where articles are divided with respect to the four previously defined collaboration classes (“Single author”, “DEAL collaboration”, “National collaboration” and “International collaboration”) is shown in Figure 9. Overall citation patterns over time are at least partially reflective of publishing patterns (Figure 4), e.g. the lowest number of citations are found in single-author and national collaborations, where the number of publications is also lowest. However, we also observe some differences between groups that would appear to be decoupled from publication volume

alone - for example, articles published by single authors have a lower proportion of articles that cite at least a single Elsevier article (38%, aggregated over all years) compared to articles published from DEAL collaborations (56%), national collaborations (65%) or international collaborations (62%). Interestingly, in terms of the total proportion of citations made to Elsevier articles, all four groups are relatively closer, with 22% of citations from single authors, 25% from DEAL collaborations, 25% from national collaborations and 23% from international collaborations being made to Elsevier articles (aggregated over all years).

In terms of variation over time, in a previous section we noted the strong decrease in Elsevier's market share of single-author articles from 2012 to 2020 (from 20.9% to 14.2%), compared to the more homogeneous publication behaviour of the three collaborative groups (Figure 4), which all displayed reduced proportions of articles published in Elsevier journals from 2016 to 2020. The evolution of citing behaviour, however, does not appear to follow the same patterns - in particular for single-author papers, we observe a long-term increase in the proportion of citations made to Elsevier articles (from 20.7% in 2012 to 21.9% in 2020), opposite to the trends found in publishing behaviour for the same group. Articles published by DEAL collaborations cited proportionally more Elsevier articles year-on-year between 2012 and 2018, and subsequently less in 2019 and 2020 (including a -1.1% YOY market share change in 2020). The picture from the national and international collaboration groups is less consistent, e.g. the proportion of citations to articles in Elsevier journals from national collaborations fluctuated in several years by $>1\%$, but without a clear trend over time, and variations in international collaborations tend to be less pronounced, with the exception of an increase of 0.9% in 2018. Changes in citing behaviour do not appear to reflect similar dynamics of publishing behaviour between different collaboration groups: of all groups, only articles published by DEAL collaborations show a decrease in the proportion of Elsevier citations concomitant with access restrictions beginning in 2018, whilst in other groups we do not find strong evidence for consistent changes in citing behaviour of Elsevier articles over time.

How did citing behaviour vary with respect to OA status? In a previous section we analysed the publishing behaviour of researchers at DEAL institutions with respect to the OA status of the articles that they published. Here, we also consider the citing behaviour of researchers at DEAL institutions with respect to OA status, with the difference that OA status refers to that of the *cited* article rather than the *published* article; that is to say, we aim to determine whether researchers preferentially cite OA (or certain types of OA) articles over closed articles, and how these patterns have changed over time with respect to articles published in Elsevier journals. Elsevier access restrictions may have hindered the ability of researchers to access (and therefore cite) articles, however, OA articles would have remained accessible. In this case, a reasonable expectation would be that the proportion of citations to OA articles in Elsevier journals would have been greater after 2018 compared to those of closed articles. Main results are shown in Figure 10.

In Figure 10A, we show the proportion of articles that cite an Elsevier article, as a function of OA type of the cited article; as an example, we observe that 45.9% of all articles published by DEAL researchers in 2012

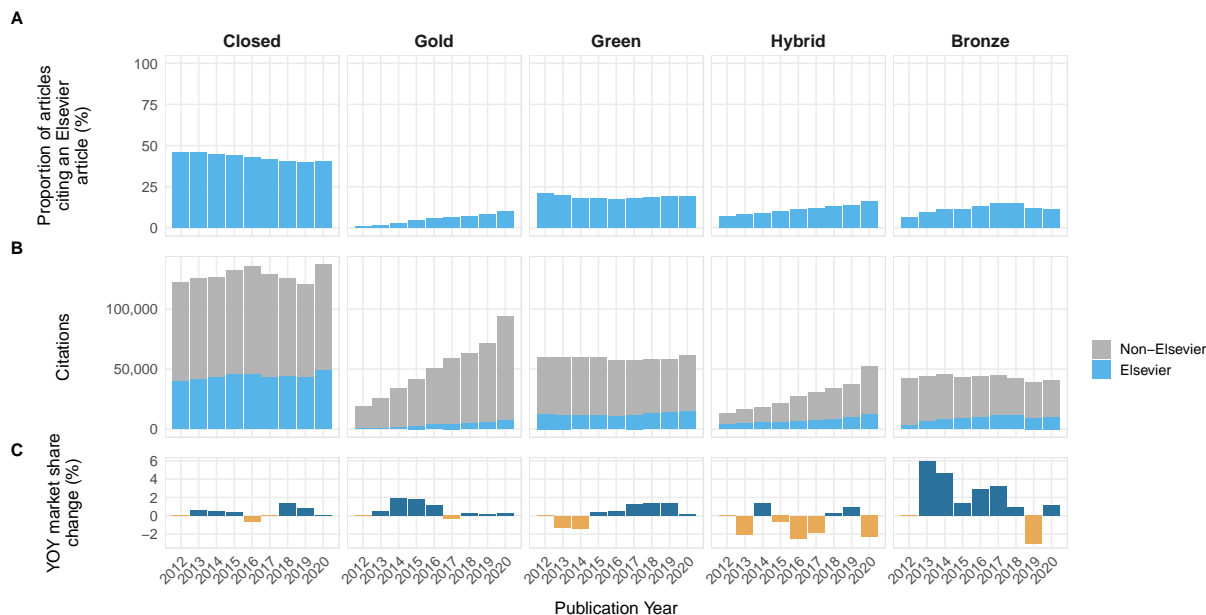


Figure 10: Citing behaviour of DEAL researchers, 2012-2020, dependent on OA status of the cited article. All data is based on citations from articles published by DEAL researchers to articles with a citation age of 2 years or less. (A) Proportion of articles published by DEAL researchers that cite at least one single Elsevier article. (B) Total number of citations from articles published DEAL researchers to articles published in Elsevier versus non-Elsevier journals. (C) Year-on-year (YOY) change in Elsevier's market share of citations from articles published by DEAL researchers.

cited at least a single Closed article in an Elsevier journal. Over time, the proportion of articles that cited a Closed article in an Elsevier journal decreased, with a final proportion of 40.1% in 2020. Comparatively, the proportion of articles that cited a Gold or Hybrid article in an Elsevier journal increased over the same time period (Gold from 0.8% to 9.9%, Hybrid from 7.3% to 16.4%). Figure 10B shows the number of references made to different types of OA between 2012 and 2020, and the proportion of which were made to articles in Elsevier journals. In general, the number of references made to Closed, Green or Bronze articles has remained relatively static over time, whilst the number of references made to Gold and Hybrid OA has rapidly increased.

Both of these results (i.e. differences in the proportion of articles citing a single Elsevier articles, and changes in the total number of citations) need to be placed in the context of general background of OA publishing trends over time. For example, whilst we observe growth in the proportion of articles citing a Gold or Hybrid OA article, the number of Gold and Hybrid OA articles has grown over the same time period (Figure 5B), and thus it is difficult to attribute such trends to true changes in the citing preference of DEAL researchers (i.e. that researchers may consciously cite more OA articles than Closed articles). However, what is clear is restricted access to Elsevier journals in 2018 did not have a disproportionate effect on the ability of researchers to cite closed articles in Elsevier journals compared to previously; in fact, after 2018 we observe YOY increases in the proportion of all citations to closed articles, that were made to articles in Elsevier journals (Figure

10C), suggesting that there were no stronger barriers to citing Elsevier articles than at other publishers.

Discussion and Conclusions

In this study, we have assessed changes in publishing and citing behaviour of researchers at DEAL institutions in Germany following Big Deal negotiations and access restrictions to Elsevier journals in 2018.

In terms of publishing behaviour, we found that Elsevier's market share of articles published by DEAL researchers has fallen from a peak of 26.3% in 2015 to 20.8% in 2020. This is in line with the hypothesis that Big Deal negotiations negatively affected researchers in their desire to publish in Elsevier journals, a trend which seems to be accelerated by the access restrictions starting in mid-2018: the largest YOY market share losses (in excess of 1% per year) appear from 2018 to 2020. We analysed these changes with respect to the timing of contract cancellations for individual DEAL institutions, research disciplines, collaboration patterns and article OA status. In broad terms, we find relatively weak evidence for any differences in publishing patterns depending on the timing of contract cancellations, or by collaboration status, although the proportion of single-author articles published in Elsevier journals has strongly decreased over the past 8 years (not limited to the period following Elsevier access restrictions). In terms of research disciplines, we observed an overall tendency towards a long-term decrease in Elsevier's market share for 19 of the 22 disciplines studied, but a high degree of variation exists over time and between individual disciplines. This may partly be the result of including a range of disciplines where publishing timelines strongly differ.

A particularly interesting feature of our results relates to that of OA publishing. Whilst OA has rapidly grown in Germany in recent years, OA growth at Elsevier has been comparatively slow - only 36.8% of all articles published by DEAL researchers in Elsevier journals in 2020 were openly available (either at the journal page, or in a Green OA repository), compared to 68.7% published in all journals. In the last 2-3 years in particular, Elsevier has lost considerable proportions of the OA market, likely driven by the successful conclusion of agreements between DEAL and other publishers (e.g. Springer Nature and Wiley; [Haucap et al., 2021](#)) and transfer of articles from DEAL researchers to those venues. Journals from the born-OA publishers Frontiers and MDPI have also gained popularity among authors affiliated with DEAL institutions in recent years (Figure 1A), which may have acted to redistribute articles from Elsevier journals to a wider range of competitors. On the other hand, the market share for Green OA increased considerably from 2016 onwards, indicating that researchers might try to make their paywalled Elsevier articles available on repositories.

In the context of previous surveys conducted in response to Elsevier access restrictions our results show a surprisingly moderate response of the publishing behaviour of DEAL researchers. Of 384 researchers surveyed at the Faculty of Medicine of the University of Münster, 29% reported that they would no longer write or review articles for Elsevier journals¹⁶, whilst 51% of respondents to a survey by in the Bibsam consortium (in response to a similar situation in Sweden) reported that they were negatively affected in their desire to

¹⁶ Archived webpage: <https://web.archive.org/web/20210429122105/https://www.uni-muenster.de/ZBMed/aktuelles/27850>

publish with Elsevier (Olson et al., 2020). In contrast, our results show only a total decrease of Elsevier's market share of articles by DEAL researchers of ~2.7% in 2019 and 2020. Continued monitoring of the situation over the following years will provide more evidence as to the long-term effect on the motivation of DEAL researchers to publish in Elsevier journals in light of continued access restrictions (or alternatively, the removal of access restrictions if future negotiations result in a successful publishing agreement between DEAL and Elsevier) and the development of new publishing options at other publishers.

In terms of citing behaviour, we found that researchers have cited proportionally less Elsevier articles after 2018 than prior to 2018, but the effect is small in comparison to that observed for publishing behaviour: overall, Elsevier's share of references (limited to “newer” references with a maximum 2-year citation age) only fell from 25% in 2018 to 24% in 2020. Such effects do not seem to imply a markedly reduced ability of researchers to cite Elsevier articles after access restrictions came into place in 2018. As with publishing behaviour, we investigated these effects with respect to the timing of contract cancellations for individual DEAL institutions, research disciplines, collaboration patterns and article OA status. However, whilst all of these units of analysis broadly agreed with the overall results, we found a lack of consistent patterns that can clearly be attributed to the implementation of Elsevier access restrictions in 2018. With respect to OA citing behaviour, where we previously noted some large-scale changes in publishing behaviour, we found that DEAL researchers are citing closed access articles in similar proportions to before the access restrictions were implemented, suggesting that no stronger barriers to citation exist for Elsevier journals compared to other publishers.

An obvious question that arises from these results, is that if researchers do not have access to Elsevier articles, why are they still able to cite them in such large volumes? A logical expectation would be that if authors are unable to access an article, they should not be able to read it and therefore not be able to cite it. This expectation is also rooted in the large number of studies over the past two decades showing a “citation advantage” of OA articles over non-OA articles (c.f. Piwowar et al., 2018), implying that accessibility and ability to cite an article are causally linked. This question may therefore be answered by considering two related mechanisms: firstly in how researchers gain access to articles, and secondly how they actually cite articles in practice.

With respect to the first mechanism, authors possess a number of strategies to access articles beyond institutional subscriptions. Sharing articles directly within informal networks of colleagues (e.g. via email) remains a permitted practice (see, for example, Elsevier's guidance page on “[Sharing and promoting your article](#)”). Some researchers have taken this a step further by requesting articles directly from their networks of followers on various social media websites (e.g. by posting an article request containing the #icanhazpdf hashtag on Twitter; Swab and Romme, 2016). Interlibrary loans (ILL), whereby a library may borrow an article from the collection of another library, and Document Delivery Services (e.g. [Subito](#)) are another option through which researchers may request articles from their institutional libraries. However, a previous analysis

on the effect of “Big Deal” cancellations has found that changes in the volume of ILL requests following such cancellations were usually small (Simard et al., 2020). Interestingly, in the previously-discussed survey of the Bibsam consortium in Sweden, 23% of respondents answered that they received access to articles through their library when denied access, yet data on article delivery services reported no increases in ILL requests for nine months following their contract cancellation (Olson et al., 2020).

Other methods that researchers may use to access articles involve so-called “shadow libraries” or “pirate OA”, where articles are made available on the web with disregard to any existing copyright. Two such venues have gained wide-scale popularity in recent years: Sci-Hub and ResearchGate. Of the respondents to the Bibsam survey, 26% answered that when they cannot access an article they instead sought access on ResearchGate, and 14% on Sci-Hub (Olson et al., 2020). A similar survey of researchers at the Faculty of Medicine of the University of Münster found that 46% of researchers sought articles on ResearchGate when access was unavailable at their university¹⁷. A previous study has found that more than 50% of articles on ResearchGate are in infringement of copyright and publishers’ policies (Jamali, 2017), which led to major publishers including Elsevier to take legal action against ResearchGate in 2018 (Else, 2018)¹⁸. Sci-Hub, founded by Alexandra Elbakyan in 2011, is another shadow library with large-scale coverage: an analysis in 2018 found that it provides access to nearly all available scholarly literature (Himmelstein et al, 2018). In Figure 11 we analysed the rates of daily Sci-Hub downloads from Germany in 2017, using freely available and geo-coded access logs that were released by Sci-Hub (data for Germany was previously aggregated by Strecker (2018)). The figure shows the proportion of all downloads that were made for Elsevier articles each day. To our knowledge, no more recent data is available beyond the end of 2017 at the time of writing this article, thus we cannot measure changes in download rates following Elsevier access restrictions in 2018; however, our data do cover the brief 6-week period that Elsevier restricted access at the beginning of 2017, during which we observe no increase in the proportion of downloads made to Elsevier articles compared to articles from other venues. Interestingly, the proportion of downloads for Elsevier articles increased dramatically by ~10% in December 2017 compared to previous months, though it is not clear what drove these increased download rates.

With respect to researchers citation practices, and how these may influence the continually high citation rates of “inaccessible” Elsevier articles, a more contentious hypothesis is that scientists do not necessarily read the full-text articles before citing them (i.e. they read only the abstracts, or do not read them at all), making the issue of access obsolete. Previous studies on the frequency and patterns of misprints in reference lists have concluded that 70-90% of references are simply copied from other articles’ reference lists (Simkin and Roychowdhury, 2005), with estimates that only ~20% of all authors that cite a paper have actually read it (Simkin and Roychowdhury, 2007). We cannot determine whether this is the case for researchers in Germany,

¹⁷Archived webpage: <https://web.archive.org/web/20210429130228/https://www.uni-muenster.de/ZBMed/aktuelles/27987>

¹⁸Legal action on the part of Elsevier and ACS resulted, for example, in the removal of around 200,000 files from ResearchGate in September 2021, see archived webpage: <https://web.archive.org/web/20211001013640/https://www.researchgate.net/blog/post/a-note-on-recent-content-takedowns>

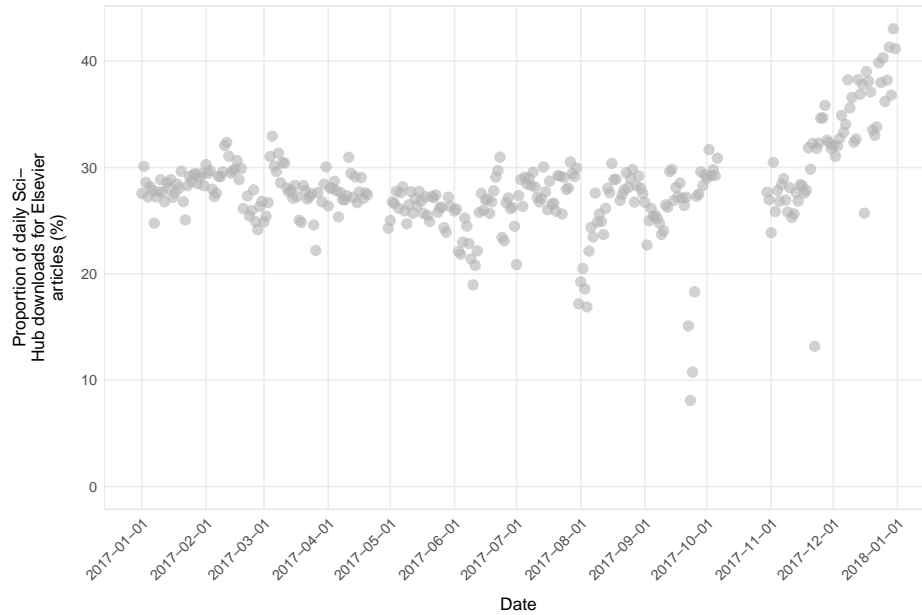


Figure 11: Proportion of daily Sci-Hub downloads from Germany in 2017 for Elsevier articles. Data from [Strecker, \(2018\)](#).

but this may present an interesting avenue for future studies that also look at differences in citation rates between OA and non-OA articles.

In summary, our aim for this study was to provide first insights into the development of researchers' publishing and citing behaviours in the aftermath of Big Deal negotiations with Elsevier and restricted access to a set of journals, from a bibliometric perspective. The results may be important for a range of stakeholders, both within and outside of Germany, involved in the negotiations of publishing agreements with commercial publishers. We hope that the methodology presented can be used as a blueprint for follow-up studies and/or long-term monitoring of these effects over a longer period of time since we assume that only recurring studies will be able to reveal the true impact of the DEAL restrictions. Our study has also shown, that - in addition to surveys that capture the perception of researchers and that rely on self-reported data - there is a strong need for the analysis of behavioural data to gain a holistic view on researchers' publication and citation processes.

Limitations and Future Directions

Our study has several limitations, which may be discussed and improved upon in future studies.

Firstly, we rely heavily on article, author and institutional data from a single bibliometric database, Dimensions. Future studies may test the robustness of our results by comparing them against results obtained through other bibliometric data sources, for example through [Web of Science](#) or [Scopus](#). These databases may have additional advantages in the quality and coverage of author role information (e.g. more complete corresponding

author information) and article types.

A second important limitation of this study is that we consider changes in publishing and citing behaviour over time with respect to publishing timelines, rather than submission timelines. Articles spend a significant proportion of time following submission in peer-review and publication cycles, and these timelines vary strongly by discipline; disciplines such as Business or Economics take 18 months on average between submission and publication (Björk and Solomon, 2013). Given that our analysis period covers at maximum a period of 30-months (i.e. to the end of December 2020) following the introduction of Elsevier access restrictions in July 2018, we are likely only capturing the early effects on researchers' behaviour. Future studies should therefore monitor these effects over longer periods; future negotiations and any potential publishing agreements made with Elsevier may also complicate such analysis further.

This study used bibliometric data to investigate potential effects of journal cancellations. However, information about the OA business model and revenue streams become more and more important in order to assess national licensing activities of libraries aiming for OA. In the case of Elsevier, it is particularly interesting to know whether authors can make use of grants to pay for publication fees. Another question is whether the open access publication would be considered as compliant by the DEAL consortium, because Elsevier publishes OA journals, which publish proceedings and are not listed in the Directory of Open Access Journals (DOAJ). Likewise, a considerable proportion of articles are made available through Elsevier's Open Archive Program after an embargo period. Future studies could make use of journal lists from related agreements between national consortia and Elsevier, as well as publisher-provided invoicing data (Jahn et al., 2021), for an in-depth analysis of OA publication activities and potential funding sources.

Moreover, this study looks at publication and citation numbers associated with a single publisher, Elsevier, and compares that development to the overall development of the total publishing market associated with German research institutions, not accounting for specific publisher profiles, changes due to market dynamics or international developments. Future studies dealing, for example, with a more detailed comparison with other major publishers or analyses how the publishing behavior of authors affiliated with DEAL institutions compares to publication and citation numbers in other countries, that were not affected by negotiations and access restrictions, will provide deeper insights into the mechanisms leading to the observed changes in Elsevier's publication and citation shares.

Finally, in this study we have taken a descriptive, quantitative approach to understanding changes in researchers publishing and citing behaviour over time. We rely solely on large-scale bibliometric data, and have not conducted any statistical modelling (e.g. to test for the interrelation between various factors that may influence publishing or citing rates) nor attempted to qualify these findings through other more qualitative methods (e.g. surveys, interviews with researchers, librarians and other stakeholders) that would be needed to understand the exact underlying mechanisms driving these changes. Authors' publishing choices are surely much more complex than what we provide as possible explanations for the bibliometric observations in this

study. It is particularly interesting, that we were able to detect a strong contradiction between self-reported data and bibliometric data on the effect of DEAL on publishing and citing. So, we encourage such follow-up studies, particularly those that use a multi-method approach to explore researchers' knowledge of Elsevier access restrictions, its effect on their day-to-day research activities, and their associated motivations for publishing (or not publishing) in Elsevier journals in future.

References

- Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4), 914–923. <https://doi.org/10.1016/j.joi.2013.09.001>
- Bornmann, L. (2018). Field classification of publications in Dimensions: A first case study testing its reliability and validity. *Scientometrics*, 117(1), 637–640. <https://doi.org/10.1007/s11192-018-2855-y>
- Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2), 151–155. <https://doi.org/10.1087/20150211>
- Chamberlain, S., Zhu, H., Jahn, N., Boettiger, C., & Ram, K. (2020). *Rcrossref: Client for Various 'CrossRef' APIs*. <https://CRAN.R-project.org/package=rcrossref>
- Crossref. (2021). *January 2021 Public Data File from Crossref*. <https://doi.org/10.13003/GU3DQMJV4>
- Else, H. (2018a). Dutch publishing giant cuts off researchers in Germany and Sweden. *Nature*. <https://doi.org/10.1038/d41586-018-05754-1>
- Else, H. (2018b). Major publishers sue ResearchGate over copyright infringement. *Nature*. <https://doi.org/10.1038/d41586-018-06945-6>
- Haucap, J., Moshgbar, N., & Wolfgang Benedikt, S. (2021). The Impact of the German 'DEAL' on Competition in the Academic Publishing Market. *CESifo, Working Paper No. 8963*. <https://www.cesifo.org/en/publikationen/2021/working-paper/impact-german-deal-competition-academic-publishing-market>
- Herzog, C., & Lunn, B. K. (2018). Response to the letter 'Field classification of publications in Dimensions: A first case study testing its reliability and validity'. *Scientometrics*, 117(1), 641–645. <https://doi.org/10.1007/s11192-018-2854-z>
- Heyman, T., Moors, P., & Storms, G. (2016). On the Cost of Knowledge: Evaluating the Boycott against Elsevier. *Frontiers in Research Metrics and Analytics*, 1. <https://doi.org/10.3389/frma.2016.00007>
- Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife*, 7, e32822. <https://doi.org/10.7554/eLife.32822>
- Hobert, A., Haupka, N., & Najko, N. (2021). Entwicklung und Typologie des Datendienstes Unpaywall. *BIBLIOTHEK – Forschung Und Praxis*. <https://doi.org/10.18452/22728>
- Hobert, A., Jahn, N., Mayr, P., Schmidt, B., & Taubert, N. (2021). Open access uptake in Germany 2010–2018: Adoption in a diverse research landscape. *Scientometrics*. <https://doi.org/10.1007/s11192-021-04002-0>
- Jahn, N., Matthias, L., & Laakso, M. (2022). Toward transparency of hybrid open access through publisher-

- provided metadata: An article-level study of Elsevier. *Journal of the Association for Information Science and Technology*, 73(1), 104–118. <https://doi.org/https://doi.org/10.1002/asi.24549>
- Jamali, H. R. (2017). Copyright compliance and infringement in ResearchGate full-text journal articles. *Scientometrics*, 112(1), 241–254. <https://doi.org/10.1007/s11192-017-2291-4>
- Kieselbach, S. (2020). *Projekt DEAL – Springer Nature Publish and Read Agreement*. <https://doi.org/10.17617/2.3174351>
- Kwiek, M. (2020). What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. *Studies in Higher Education*, 1–21. <https://doi.org/10.1080/03075079.2020.1749254>
- Milojević, S. (2012). How Are Academic Age, Productivity and Collaboration Related to Citing Behavior of Researchers? *PLoS ONE*, 7(11), e49176. <https://doi.org/10.1371/journal.pone.0049176>
- Olsson, L., Lindelöw, C. H., Österlund, L., & Jakobsson, F. (2020). Cancelling with the world's largest scholarly publisher: Lessons from the Swedish experience of having no access to Elsevier. *Insights the UKSG Journal*, 33, 13. <https://doi.org/10.1629/uksg.507>
- Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, 9(4), 734–745. <https://doi.org/10.1016/j.joi.2015.07.006>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- R Special Interest Group on Databases (R-SIG-DB), Wickham, H., & Müller, K. (2021). *DBI: R Database Interface*. <https://CRAN.R-project.org/package=DBI>
- Sánchez-Gil, S., Gorraiz, J., & Melero-Fuentes, D. (2018). Reference density trends in the major disciplines. *Journal of Informetrics*, 12(1), 42–58. <https://doi.org/10.1016/j.joi.2017.11.003>
- Sander, F., Hermann, G., Hippler, H., Meijer, G., & Schimmer, R. (2019). *Projekt DEAL – John Wiley & Son Publish and Read Agreement*. <https://doi.org/10.17617/2.3027595>
- Simard, M.-A., Priem, J., & Piwowar, H. (2020). The aftermath of Big Deal cancellations and their impact on interlibrary loans. *arXiv:2009.04287 [Cs]*. <http://arxiv.org/abs/2009.04287>
- Simkin, M. V., & Roychowdhury, V. P. (2005). Stochastic modeling of citation slips. *Scientometrics*, 62(3), 367–384. <https://doi.org/10.1007/s11192-005-0028-2>
- Simkin, Mikhail V., & Roychowdhury, V. P. (2007). A mathematical theory of citing. *Journal of the American Society for Information Science and Technology*, 58(11), 1661–1673. <https://doi.org/10.1002/asi.20653>
- Strecker, D. (2018). *Sci-Hub Downloads From Germany*. Zenodo. <https://doi.org/10.5281/ZENODO.1286284>
- Swab, M., & Romme, K. (2016). Scholarly Sharing via Twitter: #Icanhazpdf Requests for Health Sciences Literature. *Journal of the Canadian Health Libraries Association*, 37(1). <https://doi.org/10.5596/c16-009>
- Vogel, G. (2017a). German researchers start 2017 without Elsevier journals. *Science*, 355(6320), 17–17. <https://doi.org/10.1126/science.355.6320.17>
- Vogel, G. (2017b). Elsevier journals are back online at 60 German institutions that had lost access. *Science*.

<https://doi.org/10.1126/science.aal0753>

Vogel, G. (2017c). German researchers resign from Elsevier journals in push for nationwide open access.

Science. <https://doi.org/10.1126/science.aar2142>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., & Bryan, J. (2020). *Bigrquery: An Interface to Google's 'BigQuery' API*. <https://CRAN.R-project.org/package=bigrquery>