https://doi.org/10.51408/issi2025_115

Structural and Institutional Determinants of Open Access Publishing: A Macro-Perspective

Roberto Cruz Romero[1], Stephan Stahlschmidt[2]

[1]*cruzromero@dzhw.eu*
German Centre for Higher Education Research and Science Studies (DZHW), Berlin (Germany)

[2]*stahlschmidt@dzhw.eu*
German Centre for Higher Education Research and Science Studies (DZHW), Berlin (Germany)
Unit of Computational Humanities and Social Sciences (U-CHASS), EC3 Research Group, University of Granada, Granada (Spain)

Abstract

The Open Access (OA) transformation is a central component of the Open Science endeavour and is frequently addressed bibliometrically due to its engagement with publication data. Whilst most studies predominantly investigate the intra-scientific effects of the OA publication model or the framework conditions for increasing the publication rate (i.e., a responsive OA environment), fewer research is focused on the economic implications (causes and consequences) of OA for the entire science system. Thus, open science and innovation systems – as material reflections or OA consolidation – represent contexts co-determined by precise political frameworks, where policies and mandates can have direct impacts on individual challenges with clear societal implications, such as free access to scientific literature. In line with this pivotal role for knowledge transfer, the political level is a fundamental dimension with which to assess the OA patterns and the subsequent adoption of broader open science policies. Hence, both academic and regulatory dimensions of scientific production find themselves at the same crossroads, highlighting the institutional and systemic roles of that both funders and researchers play. Methodologically, we approach the relationship between policy frameworks and open science from the access dimension. We use data from three major bibliometric databases: WoS, Scopus, and OpenAlex. We are interested in observing the regional distinctions for the OA trend and try to identify geographically bound tendencies in the OA publication landscape. For that reason, we further match country-level administrative data with more specific "academic space" indicators, thus trying to uncover structural conditions that hinder or promote OA adoption. In line with recent explorations, we find that OA shows a stagnating pattern, whilst "closed" research has seen an uptake. OA costs appear flexible for richer countries than for lower income countries, which depend on a larger extent on fee-waver programs for access to read and publish in OA journals. We extend the analysis to an inferential approach through a nested logit regression, a type of multinomial logistic model to observe the probability of choosing to publish in open access compared to closed access. We discuss policy implications for the publishing landscape, as well as for the innovation-oriented scientific system.

Introduction

The Open Access (OA) transformation is a central component of the Open Science endeavour and is frequently addressed bibliometrically due to its engagement with publication data. A host of studies predominantly investigate the intra-scientific effects of the OA publication model or the framework conditions for increasing the publication rate; more broadly, studies focus on a responsive OA environment within the scientific system. Economic implications of the OA model are predominantly discussed concerning the market power of the five largest academic publishers

(Larivière, Haustein, & Mongeon, 2015), while financially weaker actors outside the scientific system, such as small and medium-sized enterprises with research interests, are initially excluded from the reuse of scientific content due to high subscription costs (Bryan & Ozcan, 2021). The latter dynamic creates a centrifugal tendency that pushes academic outputs into the realm of commodification through price-setting and access gatekeeping.

Thus, open science and innovation systems represent contexts co-determined by precise political frameworks, where policies and mandates can have direct impacts on individual challenges with clear societal implications, such as free access to scientific literature. Accordingly, the observation of the various political measures to promote the OA transformation appears practice relevant. From a political perspective, openness correspondingly refers to the conditions necessary for creating innovation incentives. In line with this pivotal role for knowledge transfer, the political level is a fundamental dimension with which to assess the OA patterns and the subsequent adoption of broader open science policies.

Nevertheless, the open knowledge generation and diffusion has to be supported by clear policy goals that complement and expand the scholarly and economic systems. As stated by Bai (2014), and reiterated by Sá and Grieco (2016), OA to research outputs needs an institutional backing so as to effectively link diverse productive actors and generate virtuous systems of research and development, as well as innovation.

Hence, OA – as a part of the open science and open innovation systems – can effectively have direct impacts on problems with great societal ramifications whilst, that is, being driven by precise policy frameworks. For example, Sá and Grieco (2016) present a still reverberating discussion about the role of open data as a result, but also as a driving force for policymaking, promoting transparency and accountability of the research output. In this case both outputs and primary data constitute the base upon which policy debates are conducted, contrasting academic, administrative, and economic perspectives.

The Institutional Perspective on Open Access

Open access has become the beacon of hope for many of its advocates. The *unpaywalled* access[1] to research should benefit scholars in disadvantageous economic circumstances (Knöchelmann, 2021), as costs for reading are eliminated through the offsetting of these by, mostly, author-sided costs. The system of OA publishing then has become dependent on large-scale agreements between publishers and research institutions (or the funding bodies supporting these). The largest European nations, for example, committed to adopting open science and OA publishing practices since the Budapest and Berlin declarations.[2]

[1] Here unpaywalled refers only to the cost-free (subscription or pay-to-read) access to scholarly research, and not to the platform with the same name.

[2] Further initiatives have been developed regarding access to publicly-funded research, such as the Helsinki Initiative for Multilingualism (Federation Of Finnish Learned Societies, Information, Publishing, Universities Norway, & European Network For Research Evaluation In The Social Sciences And The Humanities, 2019), the Vienna Principles for Scholarly Communication (Kraker

This evolving dynamic has led to the development of consequent policy frameworks, such as the European Commission's official endorsement of the San Francisco Declaration on Research Assessment (DORA), or the German Research Foundation's positioning regarding academic publishing, and as of late, the push to consolidate the Coalition for Advancing Research Assessment (CoARA). Alongside the institutional consolidation of the OA narrative, the publishing landscape has grown at a much faster pace, leaving funding bodies with many challenges to the guidelines and positions they have vis-a-vis research output. For instance, the dispersion of the OA modalities is the most telling sign of a rapidly changing environment. Beall characterises the OA publishing movement as something "concerned more with the destruction of existing institutions than with the construction of new and better ones" (2015).

This destruction is conceived within the scope of the licenses used to characterise OA (which are commonly colour-coded) and differentiate between direct (e.g., the *gold* format) and indirect routes. The latter, i.e., the *green* model, allows authors to make their works available before, during, or after the journal publication, mainly through personal or institutional repositories.[3] The *bronze* route is every sense like the green, but with a key difference regarding rights and permissions, where the bronze option can be imprecise. Transformative agreements (such as DEAL) are part of a wider scope of publishing mechanisms in which authors can choose to pay so-called article processing charges (APCs) in order to "open" their research in, predominantly, non-OA journals.

Data and Methods

To approach the relationship between policy frameworks and open science, we look firstly at the access dimension. For that, we focus on the open access information available in the bibliometric data infrastructure of the German Bibliometrics Competence Network (KB - in German). As seen in the previous plot, to frame our approach of OA growth and stagnation, we downloaded data from three major bibliographic databases: WoS, Scopus, and OpenAlex. We compare various snapshots and highlight the need for a systematised and complete dataset (see Figure 1). However, we base our exploration on the August 2024 snapshot of the OpenAlex database. We must note that bibliographic data suffers from a time-sensitive correction (as seen in Figure 1). This dynamic introduces some level of imprecision in the data exploration as for the overall counts, which we want to make noted.

Our focus is only on *articles* published between 2014 and 2023, for a time series subset of ten years. Moreover, we have two main characterisations regarding the OA information, a) as open vs. closed, and b) sub-divided into the colour categorisation

et al., 2016), the Jussieu Call for Open Science and Bibliodiversity ('Jussieu Call for Open Science and Bibliodiversity', 2017) and, more recently, the Barcelona Declaration on Open Research Infrastructure (Barcelona Declaration on Open Research Information, Kramer, Neylon, & Waltman, 2024).

[3] Specific right allocations and permissions are dependent on the editorial rights used by each publisher, and are commonly associated to a Creative Commons license (CC. See, e.g.: https://creativecommons.org/licenses/by-sa/4.0/deed.en.

described above. We delve deeper into the publisher dimension of the OA landscape and characterise two distinct dimensions of publishers: size and OA distribution. The former approach is determined based on the methodology proposed by Stephen and Stahlschmidt (2022), in which publishers are categorised according to their yearly outputs, whether in terms of articles or journals published. Seen in Figure 1, the OA momentum has been systematically driven by these larger publishers, so it becomes relevant to identify their overall weight.

Then, we are interested in observing the regional distinctions for the OA drive and the geographically bound trends in the publication landscape. So, we use of the World Bank's open data repository, and its classification in both regions and income groups. The former is based strictly on geographic bases, and hence the groupings observed. As hinted, we also look at the income level differences provided by the same dataset, dividing countries between high, upper-middle, lower-middle, and low income.

A further level of analysis is focused on the academic spaces of countries in both income and regional classifications. For this approach, we make use of the Varieties of Democracy's (V-Dem) *Academic Space* indicators, conceptualised as proxies for academic freedom (Coppedge et al., 2024). We look here into four specific dimensions of this set of measurements: freedom to research and to teach, to exchange and disseminate, to act as critics, and the institutional autonomy (in relation to the latter, the extent to which institutional autonomy is granted or, on the contrary, hindered). To fully characterise the academic space, also from an economic perspective that relates to the income and regional classifications, we complement with data on the share of public spending directed at research and development activities. Finally, in the context of academic space, we look at the Research4Life initiative, and its classification of countries that are eligible for special funding for access to specific publisher portals. This classification is highly correlated with both a geographic and income level typology, so we explore their relation further. We run logistic regressions on the publication trends in OA for an unbalanced panel of countries in the 2014-2023 timeframe using the `mlogit` package (Croissant, 2020) in the R framework.

The following results are meant as guiding insights for a deeper discussion on the role of OA publishing regarding open science and open innovations. We present preliminary data explorations that allow us to build a comprehensive perspective on, firstly, the apparent stagnation of OA and, secondly, the structural determinants that characterise it. Unless stated otherwise, all data is presented in the time frame mentioned and processed from the OpenAlex snapshot, always focusing on the access dimension.

Results

The preliminary results of the data analysis shown in the subsequent sections are an exploration of the distinct features that we seek to emphasise. Thus, the insights drawn from these approximations serve as an entrance point to a larger analytical framing and should serve as preview for the subsequent inferential analysis.

Open Access Development

Figure 1 displays the relative counts of OA publications in each of the databases considered. As seen, the trends in all three follow a most similar direction, i.e., a constant growth of OA over time (with steeper increments in OpenAlex), yet a clear slowdown (even stagnation) towards the last two periods (2022 and 2023). This behaviour illustrates our interest in the structural and institutional determinants of OA publishing, given that they constitute an underlying condition for individual researchers to opt for this publication path – that is, in addition or despite institutional policies towards closed access publications.

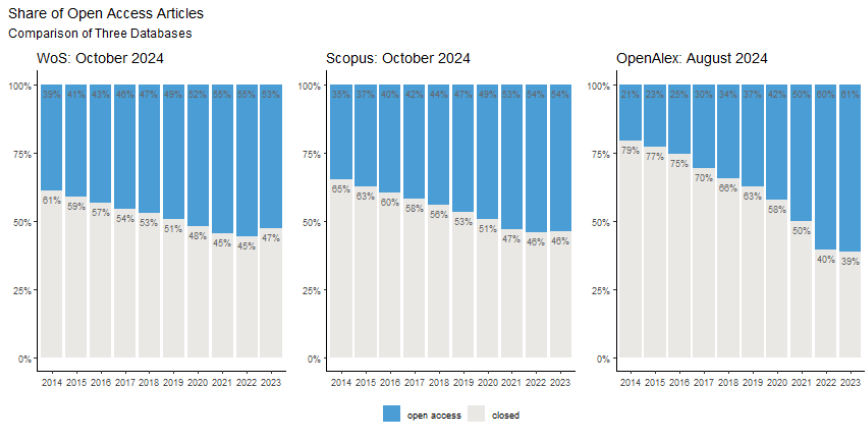

Figure 1. Open Access Trend Comparison in Web of Science, Scopus and OpenAlex (2014-2023).

Figure 2 presents a more detailed outlook of the OA trends in relation to the colour coding commonly used to indicate the corresponding licensing arrangements (see Beall, 2015).The first striking detail is the 5-p.p. change in the diamond OA articles published during the period of analysis, which reflects a limited uptake of this form of publishing standard. For clarification, *diamond* refers to those articles which licensing not only involves a creative commons copyright (CC), but crucially eliminates any payment from either side (authors and readers). This can be understood as the truly open standard. All other colour (licensing) schemes involve some other form of payment and/or limitation on the free availability of manuscripts (or data).

Looking to complementarily classify more structural dynamics that shape the allocation of resources for research, we use data from the WB that allow us to match countries to their respective income level (Figure 3). This typology is based on general thresholds of gross domestic product (GDP) that distinguish four groups: 1) High income (HI), 2) upper middle income (UMI), 3) lower middle income (LMI), and 4) lower income (LI) countries. Methodologically, we note that not all countries listed on the WB data are present in the bibliographic data from OpenAlex. Presumably, not every country that is listed had an academic affiliation which produced an article included in the database (a double contingency that limits the

scope of this study). We complement this perspective with academic space indicators.

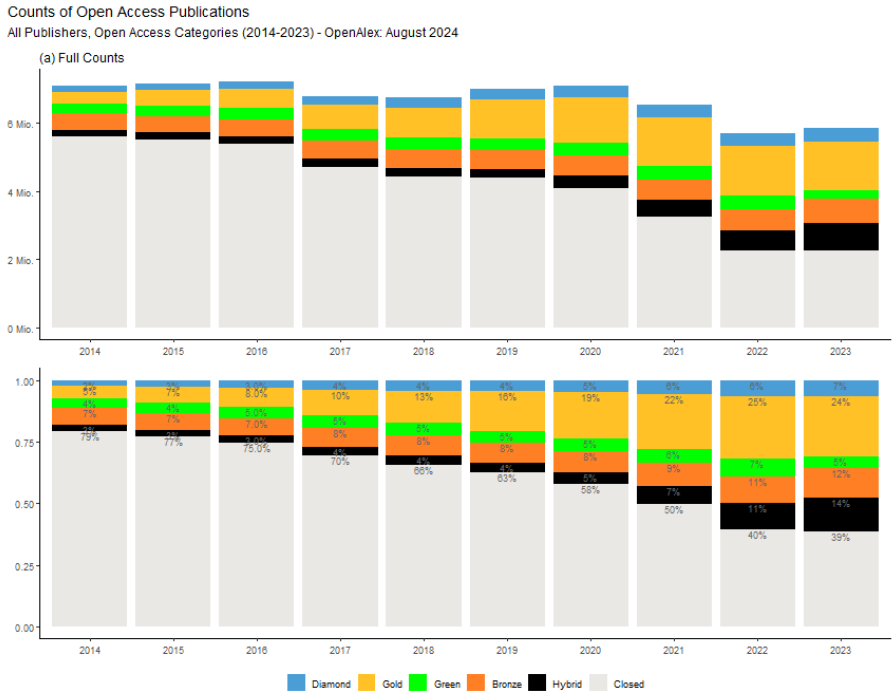

Figure 2. Open Access Trend Breakdown in OpenAlex (2014-2023).

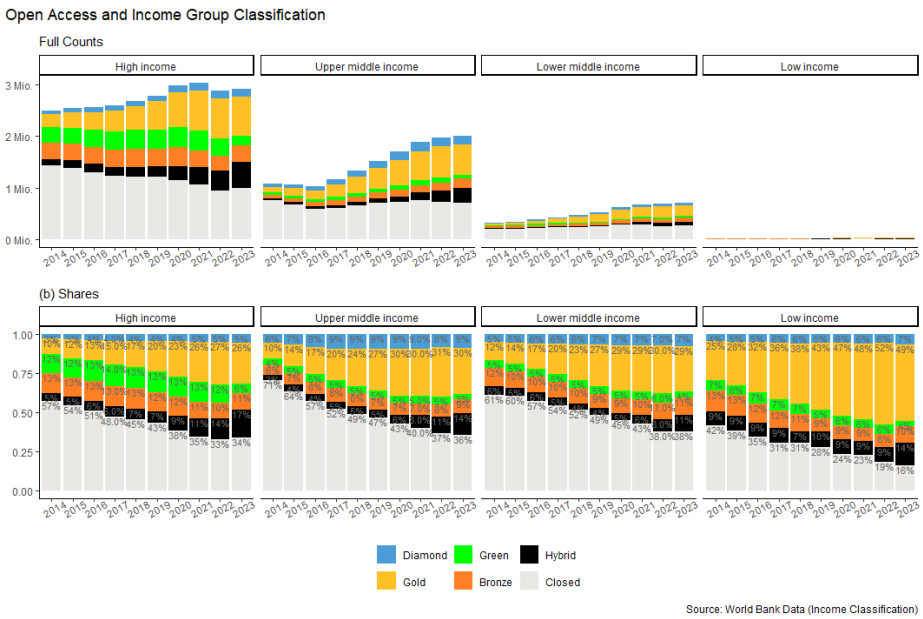

Figure 3. Open Access Breakdown by Income Classification (2014-2023).

To that end, we use the Varieties of Democracy (V-Dem) project's ***Academic Freedom Index***, or AFi. This index is composed of a series of disaggregated indicators that make up the "academic space", which characterises more thoroughly the *de jure* and *de facto* gaps in expert based and collected factual data (see Spannagel, Kinzelbach, & Saliba, 2020). We note that this index is not flawless and recognise the methodological shortcomings this type of construct may imply. We do not make use of the AFi as such, but rather we look at four precise indicators that make up the academic space characterisation included in the V-Dem dataset.[4] The four dimensions we consider are: 1) Research and teaching, 2) academic exchange and dissemination, 3) academics as critics, and 4) institutional autonomy.

Finally, we match the data with the R4L initiative dataset. The organisation funds and subsidises access to scholarly content by cataloguing countries according to their scores in different indices that relate to human development. Countries are then placed in eligibility groups to provide "institutions in low-and middle-income countries with online access to academic and professional peer-reviewed content" (see https://www.research4life.org/about/). In this sense, the lists of the R4L initiative are a proxy of reduced access to research outputs and, therefore, of analytical relevance to match with the WB and OpenAlex data.

Logistic Regression

Looking to tie together the analytical elements hitherto discussed, we now turn to an inferential analysis based on a multinomial logistic regression (MLR). We approach the analysis from this perspective given that the response variable we are interested in has the characteristic of being nominal (i.e., neither ordinal nor numeric), and we are focused on the probabilistic changes from one category to the other. In this sense, we take the bibliometric data grouped at the country level (based on author affiliation),[5] and count the total number of publications, as well as the distribution of these in OA categories – our base category, however, is closed access. To include all necessary features of interest, we match the data with the previous dataset already presented and discussed, i.e., with economic and administrative data from the WB (GDP and income groups), from V-Dem (AFi indicators), and from R4L initiative's group classification.

The data has an unbalanced panel structure, since we have the timeseries of ten years (2014-2023) with distinct number of observations in each of the categories of OA, our response variable. We group by country and year so as to aggregate the data to a macro-level of analysis.

The crux of the modelling structure, however, lies at the nature of the conditions for estimating a multinomial regression. Since we are dealing with alternatives of similar conditions, i.e., with most of the options that could be grouped under the larger nest of OA, we recur to the implementation of a nested logit model. To that end, we recode the data to, firstly, differentiate between the nest categories, i.e., open v.

[4] We use version 14 of the country/year dataset, in which the academic space indicators are found in section 3.15.4 of the Codebook.

[5] We do not estimate fractional counting when assigning country affiliation distributions. For matters of exploration and direct interpretation, we proceed with full counting of authors.

closed access. Then, we estimate the change between nested sub-categories (diamond, gold, green, bronze, and hybrid), since we part from the characteristic that "some alternatives may be joined in several groups" (Croissant, 2020, p. 21). Following this rationale, the nested logit model relaxes a key assumption regarding the cross-elasticity of alternatives (this is, that "the introduction of any new mode or the improvement of any existing mode will affect all other modes proportionally" (Forinash & Koppelman, 1993, p. 98) – with modes referring here to the alternatives). In this sense, the condition of independence of irrelevant alternatives (IIA) is relaxed and allows for correlated error terms of the groups of alternatives; put otherwise, this estimation technique parts from the premise that not only case-specific characteristics determine the choice, but also alternative-specific characteristics have a probabilistic weight for maximising the utility in the decision-making. Figure 4 (an adaptation from Forinash & Koppelman, 1993) details the nested estimation approach. As seen in the figure, we differentiate between these characteristics, viz. case (individual or, in this scenario, country-level) characteristics and alternative characteristics. The former are the same across alternatives, whilst the latter vary across choices.

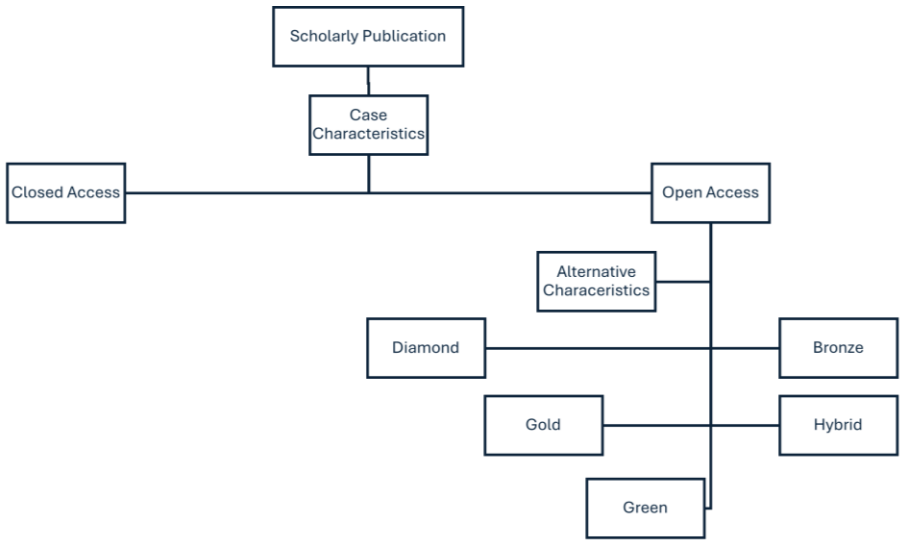

Figure 4. Example of Nested Estimation Approach.

Limitations and Next Steps

The pending analysis of the regression estimates follows a data intensive computation with the `mlogit` package in R (Croissant, 2020). The structure of the data (long format, i.e., short *T* and large *N*) creates computational demands that make it necessary to recur to a special server. The latter's capacity varies according to overall shared usage, which caused unforeseen delays in the process of preparing this submission. Hence, the inferential analysis will be further developed and presented at the conference in full scope.

Acknowledgments

This contribution is financed by the German Federal Ministry for Education and Research (BMBF) in its Open Access Culture funding line (project number: 16KOA014).

References

Alatas, S. F. (2003). Academic Dependency and the Global Division of Labour in the Social Sciences. *Current Sociology*, *51*(6), 599–613. SAGE Publications Ltd.

Bai, C. (2014). Promoting research and innovation with open access. *National Science Review*, *1*(3), 323.

Barcelona Declaration on Open Research Information, Kramer, B., Neylon, C., & Waltman, L. (2024, April 16). Barcelona Declaration on Open Research Information. Zenodo. Retrieved June 18, 2024, from https://zenodo.org/doi/10.5281/zenodo.10958522

Beall, J. (2015). What the Open-Access Movement Doesn't Want You to Know. *Academe (May-June 2015: 'I'll Tell It and Think It and Speak It and Breathe It')*, *101*(3). Retrieved June 17, 2024, from https://www.aaup.org/issue/may-june-2015-ill-tell-it-and-think-it-and-speak-it-and-breathe-it

Berger, M. (2021). Bibliodiversity at the Centre: Decolonizing Open Access. *Development and Change*, *52*(2), 383–404.

Bryan, K. A., & Ozcan, Y. (2021). The Impact of Open Access Mandates on Invention. *The Review of Economics and Statistics*, *103*(5), 954–967.

Cope, B., & Kalantzis, M. (2014). Changing knowledge ecologies and the transformation of the scholarly journal. In B. Cope & A. Phillips (Eds.), *The Future of the Academic Journal (Second Edition)* (pp. 9–83). Oxford: Chandos Publishing. Retrieved November 7, 2022, from https://www.sciencedirect.com/science/article/pii/B9781843347835500021

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Bernhard, Mi., et al. (2024). V-Dem [Country–Year/Country–Date] Dataset v14. Varieties of Democracy (V-Dem) Project. Retrieved from https://www.v-dem.net/data/the-v-dem-dataset/country-date-v-dem-v14/

Croissant, Y. (2020). Estimation of Random Utility Models in *R*: The **mlogit** Package. *Journal of Statistical Software*, *95*(11). Retrieved February 4, 2025, from http://www.jstatsoft.org/v95/i11/

Delgado-López-Cozar, E., & Martín-Martín, A. (2024). La ruta de oro de la publicación científica: Del negocio de las revistas a las revistas negocio: La fuente del negocio editorial: el negocio bibliométrico de la evaluación científica. *Revista Mediterránea de Comunicación*. Retrieved February 22, 2024, from https://www.mediterranea-comunicacion.org/article/view/26763

Federation Of Finnish Learned Societies, Information, T. C. F. P., Publishing, T. F. A. F. S., Universities Norway, & European Network For Research Evaluation In The Social Sciences And The Humanities. (2019). Helsinki Initiative on Multilingualism in Scholarly Communication, 621757 Bytes. figshare.

Forinash, C. V., & Koppelman, F. S. (1993). Application and Interpretation of Nested Logit Models of Intercity Mode Choice. *Transportation Research Record*, (1413). Retrieved February 4, 2025, from https://trid.trb.org/View/385097

Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLOS Medicine*, *15*(4), e1002549.

Jussieu Call for Open Science and Bibliodiversity. (2017). . Retrieved March 18, 2024, from https://jussieucall.org/jussieu-call/

Khanna, S., Ball, J., Alperin, J. P., & Willinsky, J. (2022). Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. *Quantitative Science Studies*, *3*(4), 912–930.

Khoo, S. Y.-S. (2019). Article Processing Charge Hyperinflation and Price Insensitivity: An Open Access Sequel to the Serials Crisis. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, *29*(1), 1–18.

Knöchelmann, M. (2021). The Democratisation Myth: Open Access and the Solidification of Epistemic Injustices. *Science & Technology Studies*, *34*(2), 65–89.

Knöchelmann, M. (2023). Herausgeberschaft und Verantwortung: Über die Un-/Abhängigkeit wissenschaftlicher Fachzeitschriften. *Bibliothek Forschung und Praxis*, *47*(2), 393–406. De Gruyter.

Kraker, P., Dörler, D., Ferus, A., Gutounig, R., Heigl, F., Kaier, C., Rieck, K., et al. (2016). The Vienna Principles: A Vision for Scholarly Communication in the 21st Century. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, *69*(3–4), 436–446.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, *10*(6), e0127502.

Leonelli, S. (2022). Open Science and Epistemic Diversity: Friends or Foes? *Philosophy of Science*, *89*(5), 991–1001.

Picarra, M. (2015). *Open Access To Scientific Information: Facilitating Knowledge Transfer And Technological Innovation From The Academic To The Private Sector* (Briefing Paper No. 611742). PASTEUR4OA (p. 9). Brussels: European Commision. Retrieved from Zenodo.

Sá, C., & Grieco, J. (2016, September). Open data for science, policy, and the public good. *Review of Policy Research*. WILEY.

Spannagel, J., Kinzelbach, K., & Saliba, I. (2020). *The Academic Freedom Index and Other New Indicators Relating to Academic Space: An Introduction* (No. 26). Users Working Paper (p. 28). Gothenburg: V-Dem Institute, University of Gothenburg. Retrieved from https://www.v-dem.net/media/publications/users_working_paper_26.pdf

Spiekermann, K. (2020). Epistemic network injustice. *Politics, Philosophy & Economics*, *19*(1), 83–101. SAGE Publications.

Van Noorden, R. (2013). Open access: The true cost of science publishing. *Nature*, *495*(7442), 426–429. Nature Publishing Group.