CEPN Centre d'économie de l'Université Paris Nord CNRS UMR n° 7234

Document de travail N° 2019-08
Axe : Macroéconomie Appliquée Finance et Mondialisation

Construction of a Normalized Open Access Indicator (NOAI)

Abdelghani Maddi,
abdelghani.maddi@hceres.fr
CEPN, Université Paris 13, Sorbonne Paris Cité

October 2019

Abstract: The issue of Open Access (OA) in research is attracting growing interest both within the scientific community and on the political scene. Some centers specializing in the production of science indicators now include OA indicators by institution. In its 2019 ranking, the Centre for Science and Technology Studies (CWTS) provides a ranking of institutions according to their share of open access publications. This gives an idea of the degree of openness of institutions. However, the fact of not taking into account the disciplinary specificities and the specialization of the institutions makes the rankings based on the shares of the OA publications biased. We show that open access publishing practices vary considerably by discipline. As a result, we propose two methods of normalization of OA share; by WoS subject categories and by OST disciplines. Normalization corrects OA's share taking into account disciplinary practices. This allows a better comparability of different actors.

Keywords: Open Access, normalisation, ranking, institution, bibliometrics

Construction of a Normalized Open Access Indicator (NOAI)

Abdelghani Maddi
abdelghani.maddi@hceres.fr
Observatoire des Sciences et Techniques, Hcéres,
2 rue Albert Einstein, Paris, 75013 France ;
Université Paris 13, Sorbonne Paris Cité, CEPN (CNRS, UMR 7234)

Abstract

The issue of Open Access (OA) in research is attracting growing interest both within the scientific community and on the political scene. Some centers specializing in the production of science indicators now include OA indicators by institution. In its 2019 ranking, the Centre for Science and Technology Studies (CWTS) provides a ranking of institutions according to their share of open access publications. This gives an idea of the degree of openness of institutions. However, the fact of not taking into account the disciplinary specificities and the specialization of the institutions makes the rankings based on the shares of the OA publications biased. We show that open access publishing practices vary considerably by discipline. As a result, we propose two methods of normalization of OA share; by WoS subject categories and by OST disciplines. Normalization corrects OA's share taking into account disciplinary practices. This allows a better comparability of different actors.

Introduction

The issue of Open Access (OA) in research is attracting growing interest both within the scientific community and on the political scene. For researchers, the OA reduces barriers to accessing research results and their use (Martín-Martín *et al.* 2018). The OA ensures a better dissemination of knowledge and contributes to accelerating the development of science. The scientific literature on the subject shows that publications in OA are much more cited than their counterparts for whom no Open Access version is available (Antelman, 2004; Harnad *et al.* 2004; Eysenbach, 2006; Piwowar *et al.* 2018). Thus, the academic impact of researchers and institutions increases as the number of OA publications increases. As a result, researchers are increasingly prone to publish in OA in order to make their results more accessible, with the prospect of a higher and faster impact (Antelman, 2017).

For funders, the stakes are different as the OA does not necessarily mean "free" and may even generate new costs (Borrego, 2016; Anderson, 2017a, 2017b). There are two main types of OA publications; "Gold" and "Green" (Björk *et al.* 2010; Björk *et al.* 2014; Björk, 2017). Both types allow readers to access the full text. Gold OA covers especially Creative

Commons licensed articles published in journals listed in the Directory of Open Access Journals (DOAJ) (Gargouri *et al.*, 2012; Archambault *et al.*, 2014; Bolick, 2017). These are journals that rely on an economic model based on "Article Processing Charges (APC)" which are fees paid by the authors (usually via their institution). The Green OA represents articles deposited in open archives. Some non-OA journals allow authors to submit either the version before the peer review (Preprint) or the post-evaluation – peer reviewed – version (Post-print). Apart from these two types, there is another category of OA called "Bronze". It includes articles published in journals that do not have a license (Creative Commons) or an unidentified OA status (which can be temporary) in databases. Their status can evolve over time to become Gold or Green OA (or both at the same time).

It is important to distinguish between the status of the publication and that of the journal. A journal can have three statuses; OA, not OA or hybrid. An OA journal publishes OA-type articles, while a hybrid journal is a closed (fee-paying) journal that gives authors the option to publish in OA for a fee (APC). The resulting publication will also be of the "Gold" OA type (Walker *et al.* 1998; Laakso *et al.* 2012, 2013; Björk, 2016a; Martín-Martín *et al.* 2018). It is possible for an article to have multiple OA statuses at the same time. For example, an article can be published in an OA Gold journal and then deposit in a national archive (Green). More generally, beyond the deposits made by the authors, the OA journals can fully dump their contents in an archive like PubMed Central. It should be noted that the reliability of the data is variable depending on the status. If the information on Gold OA can be considered reliable because relatively stable, it is not the same for Bronze status, volatile by nature. The open archives are fed continuously by the authors or by the journals, the information on Green OA is, meanwhile, also quite ephemeral (Björk, 2016a; Martín-Martín *et al.* 2018).

In addition to subscription costs that institutions must subscribe to (to ensure that their researchers have access to publications) they are now more and more led to pay the costs of publication in OA that can reach 5000 euros for one publication (Simth *et al.*, 2017; Antelman, 2017). This amounts to paying twice for OA publications. For this reason, some consider that the current system based on publisher subscriptions becomes anachronistic and it is imperative to upgrade to 100% OA. The concept of the "Big Deal" then emerged to denote the difficulty of changing the publishing market system as it works today (Schiermeier & Mega, 2017; Anderson, 2017a; Université Konstanz, 2014; Université de Montréal, 2017). Schimmer et al. (2015) shows that if the WoS only indexed articles (1.5 million in 2013), the

unit cost in the current subscription system would be 5000 euros per article (the overall cost of subscriptions is estimated at EUR 7.6 billion). While in a system that only operates according to OA rules, the community would produce 2 million articles at a unit cost of 3800 euros (with the same budget). Hence the interest of switching to 100% OA. Several countries, like the Netherlands, Germany and the United Kingdom have started negotiations with publishers to find an agreement around the "Big Deal" including subscriptions and APCs. In France, after 13 months of negotiations, in order to limit the rise in subscription prices and take into account APCs, the national consortium "Couperin.org" decides not to renew the agreement with Springer since 2018[1]. At the beginning of July 2018, under the leadership of the French Minister of Higher Education Frédérique Vidal, France set up a National Open Science[2] Plan which "makes open access mandatory for publications and data from project-based research (financed by public funds)".

Relationships remain tense between funders (governments, research organizations, etc.) and publishers who have a stranglehold on the world of scientific publishing, for several reasons. First, the publishing market operates according to the rules of an oligopolistic market. A small number of concentrated publishers who have become multinationals through mergers and acquisitions now hold the keys to the market and influence the price of subscriptions that does not depend on the confrontation of supply and demand. Their bargaining power is very strong given the need for information (access to publications) of their interlocutors. Second, publishers legally own journals; and therefore their publications whose property rights are ceded to them by the authors (Björk, 2016b).

Over the last ten years, the world of science has witnessed a rise in a discourse in the same direction as that of the funders. A large scientific community agrees that research results should be accessible not only to all researchers but also to society as a whole (Tennant *et al.* 2016). Since research is funded mainly by taxpayers, it is unjustified that publications are held exclusively by multinationals, which are demanding increasingly high fees. In addition to the immediacy of the sharing of research content, opening up science also has virtues at the global level. Due to lack of funds, some researchers in low-income countries do not have the same access to publications as their counterparts in high-income countries. Moving to a 100% OA system would provide greater equity.

[1] https://www.couperin.org/services-et-prospective/grilles-d-evaluation-ressources/261-a-la-une/1333-couperin-ne-renouvelle-pas-l-accord-national-passe-avec-springer
[2] https://www.ouvrirlascience.fr/open-science/

The world of scientific publishing remains very unstable and is witnessing rapid and profound transformations. In this context, it is imperative for policymakers and funders to have an "overview" of open science in order to streamline their decisions and guide their arbitrations regarding AO publications. As a result, they become the first seekers for OA indicators.

Some centers specializing in the production of science indicators now include OA indicators by institution. In its 2019 ranking, the Centre for Science and Technology Studies (CWTS, 2019) provides a ranking of institutions according to their share of open access publications (with all variations by type of OA). This gives an idea of the degree of openness of institutions. However, the fact of not taking into account the disciplinary specificities and the specialization of the institutions makes the rankings based on the OA's share biased. Thus, open access publishing practices vary considerably by discipline. OA's share is very high in Fundamental Biology and much less so in Computer Science and Engineering.

The purpose of this paper is double. First, draw up an overview of the OA in the WoS database in terms of volume, evolution and disciplinary distribution. Second, propose two methods of normalization of OA share; by WoS subject categories and by OST disciplines. This indicator corrects OA's share taking into account disciplinary practices. This allows a better comparability of different actors (institutions and countries).

Data & method

Since 2014, the provider of the WoS database, Carivate Analytics (CA), retrospectively identifies the status of OA publications. In 2017, CA signed a partnership with ImpactStory (https://impactstory.org/) to better identify OA's status. Moreover, the data of the WoS are enriched by the OST, particularly for the French institutions that carry out annual identification of their publications within the framework of IPERU-OST program (Scientific Production Indicators of Academic Research Institutions).

We calculate the Normalized Open Access Index (NOAI) for the first 50 producing countries and for the French institutions included in the IPERU-OST program for the year 2018. In total, 124 institutions (organizations research, universities, etc.).

Normalization is done in two stages. First, calculate the share of OA ($^{OA_{ij}}/_{x_{ij}}$) by institution (or country) and by discipline, then report it to the same share at the global level ($^{OA_{wj}}/_{X_{wj}}$).

$$OA_{S_{ij}} = \frac{^{OA_{ij}}/_{x_{ij}}}{^{OA_{wj}}/_{X_{wj}}}$$

In a second step, to have an overall OA indicator by institution / country, it is possible to calculate a weighted average by the number of publications per discipline. We then obtain the Normalized Open Access Index (NOAI):

$$NOAI_i = \frac{\sum(OA_{S_{ij}} \times x_{ij})}{x_i}$$

Descriptive statistics

The WoS database contains more than 12 million publications for which Open Access status is provided. Over the 2012-2017 period, OA publications represent 30% of the entire base. The WoS has 5,000 journals in OA (excluding hybrid journals) out of a total of 14,000.

Figure 1: share of world Open Access publications (WoS)

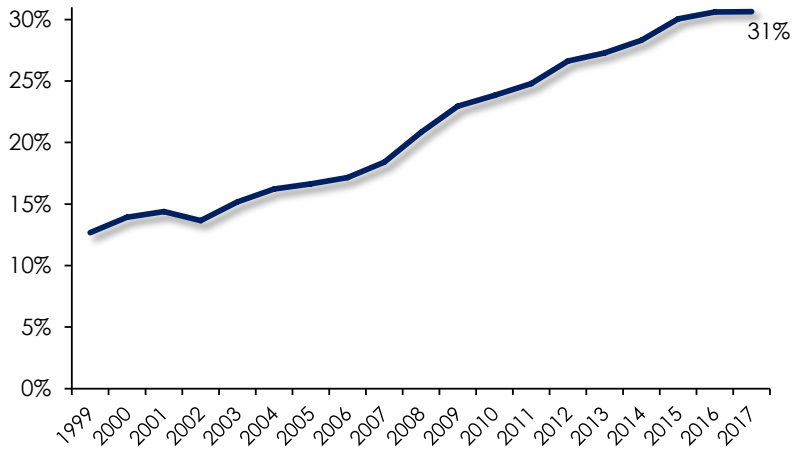


Figure 1 shows that the share of open access publications increases continuously between 2000 and 2017, reaching 31% at the end of the period, whereas the share was 14% in 2000. It should be noted that the bulk of these publications in OA is of type "Gold" (or "Bronze" which is gradually transformed into "Gold") which is a peculiarity of WoS database.

Figure 2: share of world Open Access publications by ERC panel

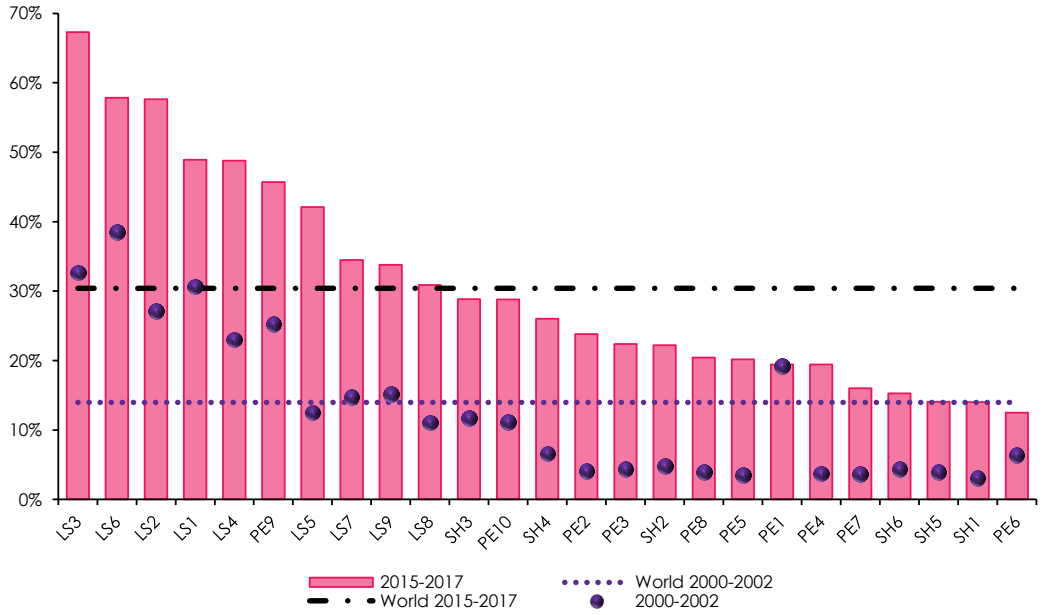
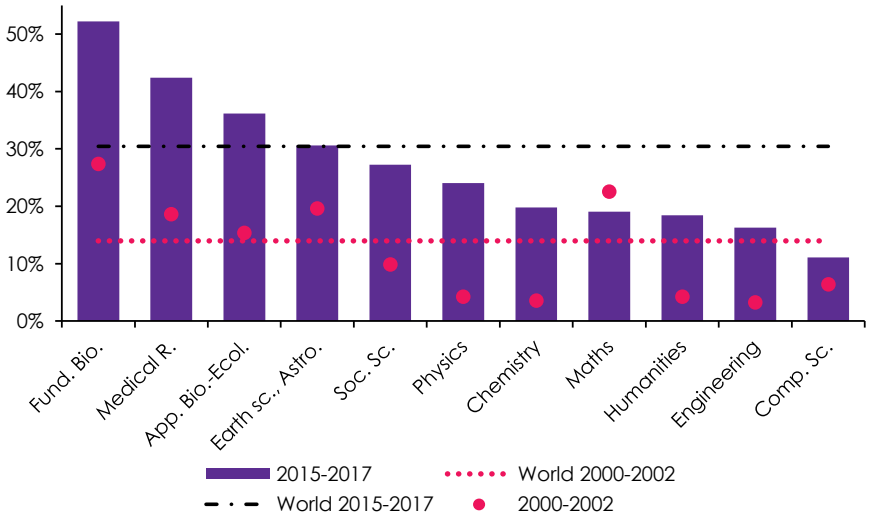


Figure 2 shows the distribution of OA publications in the ERC Panels. There is a great disparity between the different panels as to the "practices of openness". The share of OA publications has increased significantly for almost all panels.

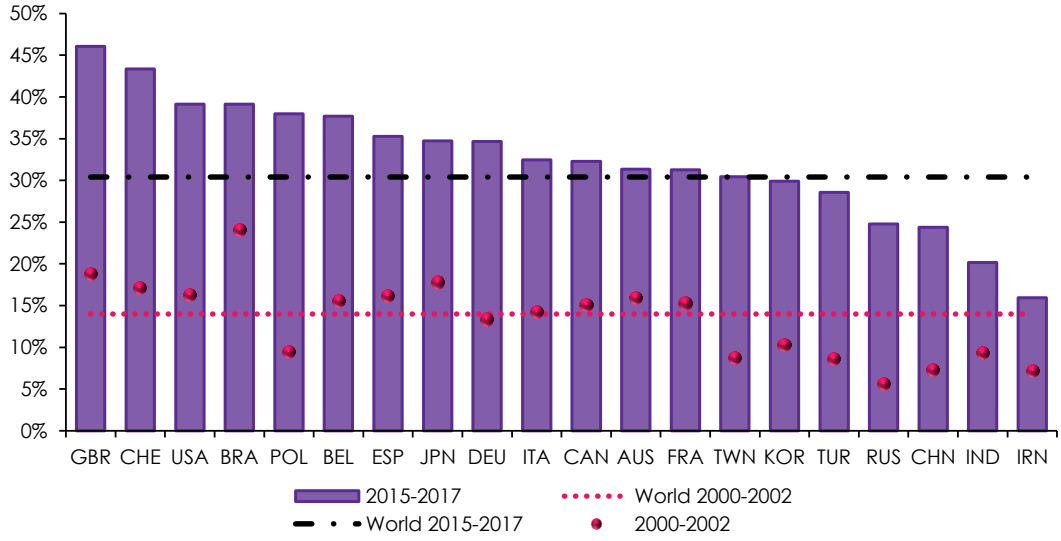
In 2015-17, the share varies between 12% for the PE6 (Computer Science and Informatics) panel and 70% for the LS3 (Cellular and Developmental Biology) panel. Overall, the share of OA is relatively high in the areas of "life sciences", and low in particular in the panels of "Social Sciences and Humanities" and "Physical Sciences and Engineering". The Universe Sciences (PE9) is the panel with the highest proportion of OA (46%) in these last two domains. The rest of the panels have a lower than world average share in both periods (2000-02 and 2015-17).

Figure 3: share of world Open Access publications by OST disciplines



With regard to the nomenclature in 11 disciplines of the OST, like the ERC panels, the proportion of OA publications is relatively high in fundamental biology (52%) and in medical research (42%) and low in humanities (18%) in engineering (16%) and computer science (11%). However, it is important to remember that these rates do not necessarily reflect the real practices of "openness" in these disciplines. For example, OA's share would be much higher in mathematics and physics if open archives such as ArXiv were taken into account. This shows that it is imperative to take into account the degree of representation of the database for this indicator and that it is essential to normalize when comparing research actors (given their disciplinary orientations).

Figure 4: share of Open Access publications, top 20 countries



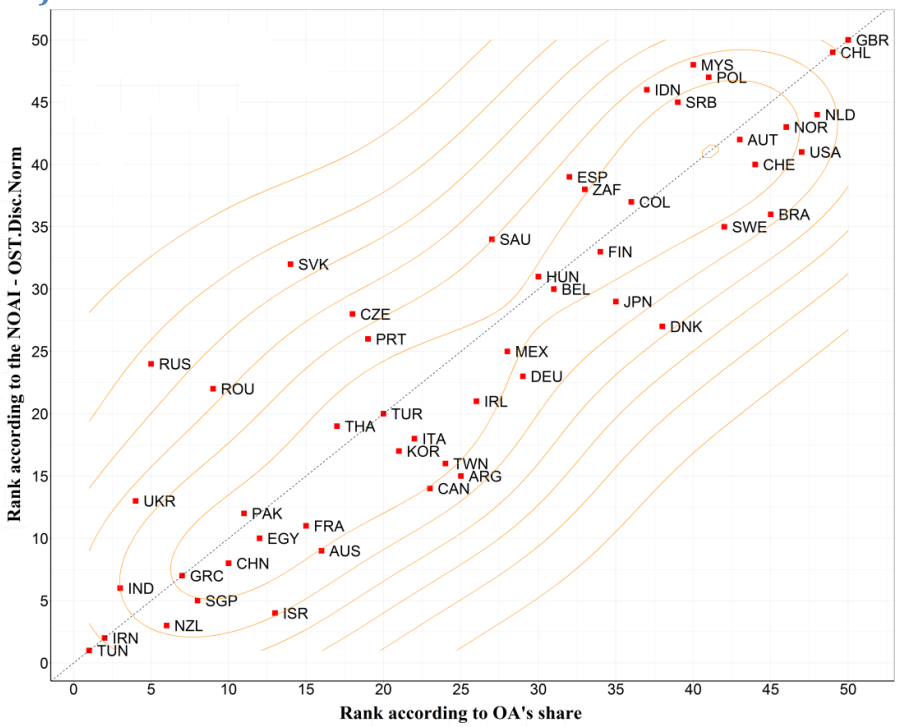
Among the top 20 producers, the share of OA publications is very contrasting. The United Kingdom is the country with the largest share (46%), 15% above the world average, followed by Switzerland (43%). The United States is followed by Brazil with a similar share of OA publications (39%). France has a slightly higher share than the world average (31%).

Figure 4 also shows that countries with a high specialization engineering, mathematics and chemistry have low OA shares, such as Russia, China, India and Iran (OST, 2019).

Application at country and institutional levels

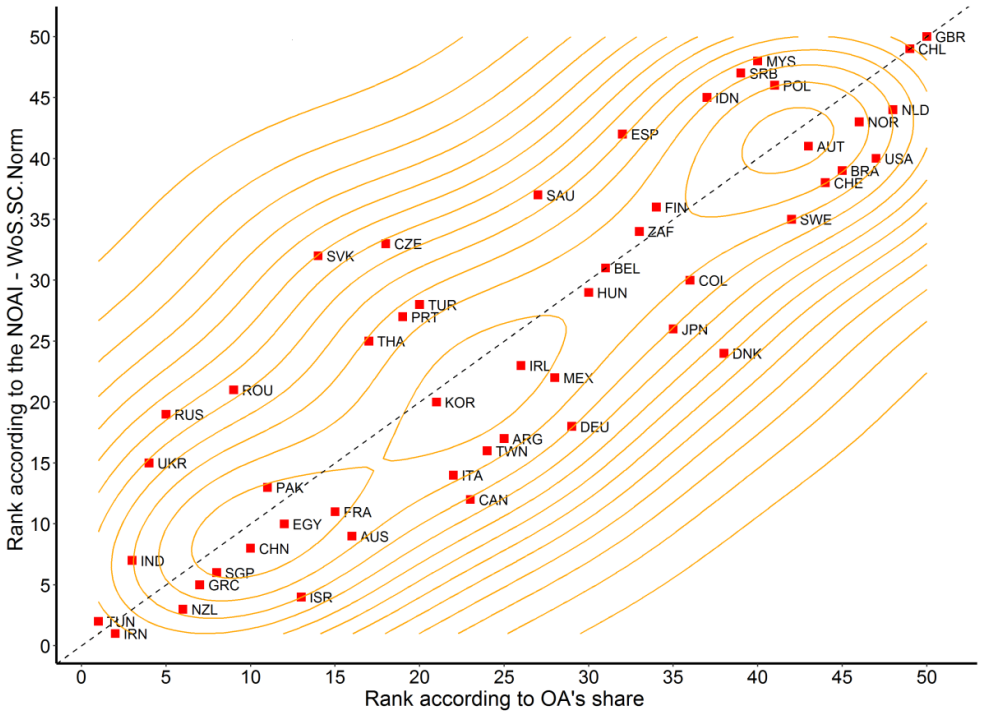
Figure 5 shows the rank of countries according to their share of open access publications (abscissa axis), and their rank according to the NOAI using normalisation at OST disciplines level (ordinate axis). The rank is in descending order. That is, the countries with the highest OA share are to the right of axis. Rank "1" represents the country with the lowest share of OA. The figure shows that globally the two ranks are correlated. However, the rank changes considerably for some countries like Russia, which gains 20 places by normalizing the proportion of OA by discipline. We also note that a number of low-income countries specializing in low-OA share disciplines are moving up in ranking with the normalization.

Figure 5: rank of countries by OA and NOAI (normalization by OST disciplines)



Although France is very specialized in mathematics, it loses some places in the ranking on the normalized indicator. This could be explained by the fact that France has a relatively more diversify disciplinary profile, unlike low-income countries with disciplinary profiles that are very much oriented towards one or two disciplines. The United Kingdom keeps its first place on both indicators.

Figure 6: rank of countries by OA and NOAI (normalization by WoS subject categories)



The normalization at the level of the 255 WoS subject categories shows some differences compared to OST disciplines based normalization (Figure 6). Some countries keep their position regardless of the method of normalization (United Kingdom, France, etc.), while others change their position like Russia losing 5 places and Turkey move up by 9. Overall the ranks remain substantially similar on both types of normalization.

Figure 7: rank of institutions by OA and NOAI (normalization by WoS subject categories)

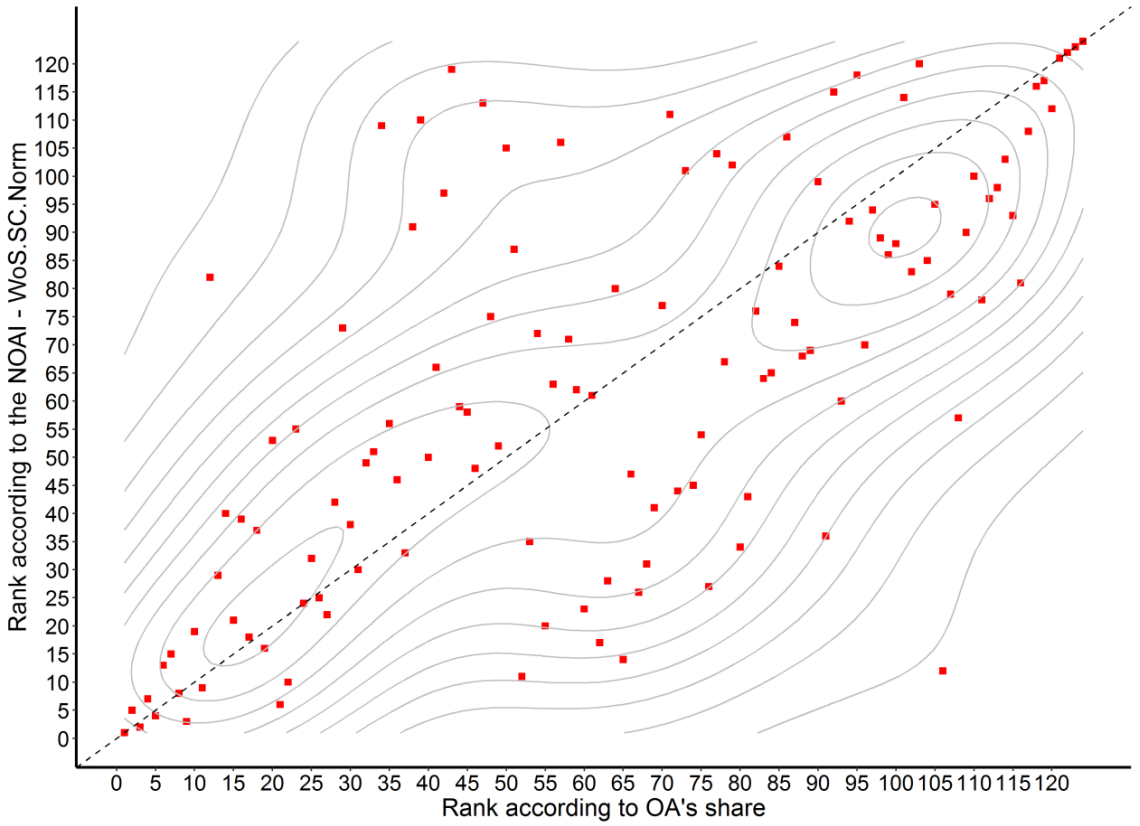


Figure 7 shows that there is a large variation in ranking according to the indicator used, in particular for certain institutions[3]. The figure shows that a good part of schools of engineering and specialized institutions in engineering, computer science or social sciences and humanities are located to the left of the bisector. That is, they gain rank after normalization of share of OA publications. In contrast, institutions to the right of the bisector are more oriented towards basic biology, applied biology-ecology, and medicine. The two rankings nevertheless remain globally correlated (the correlation coefficient is equal to 0.68).

[3] For confidentiality reasons, we cannot display the institution's names.

Figure 8: rank of institutions by OA and NOAI (normalization by OST disciplines)

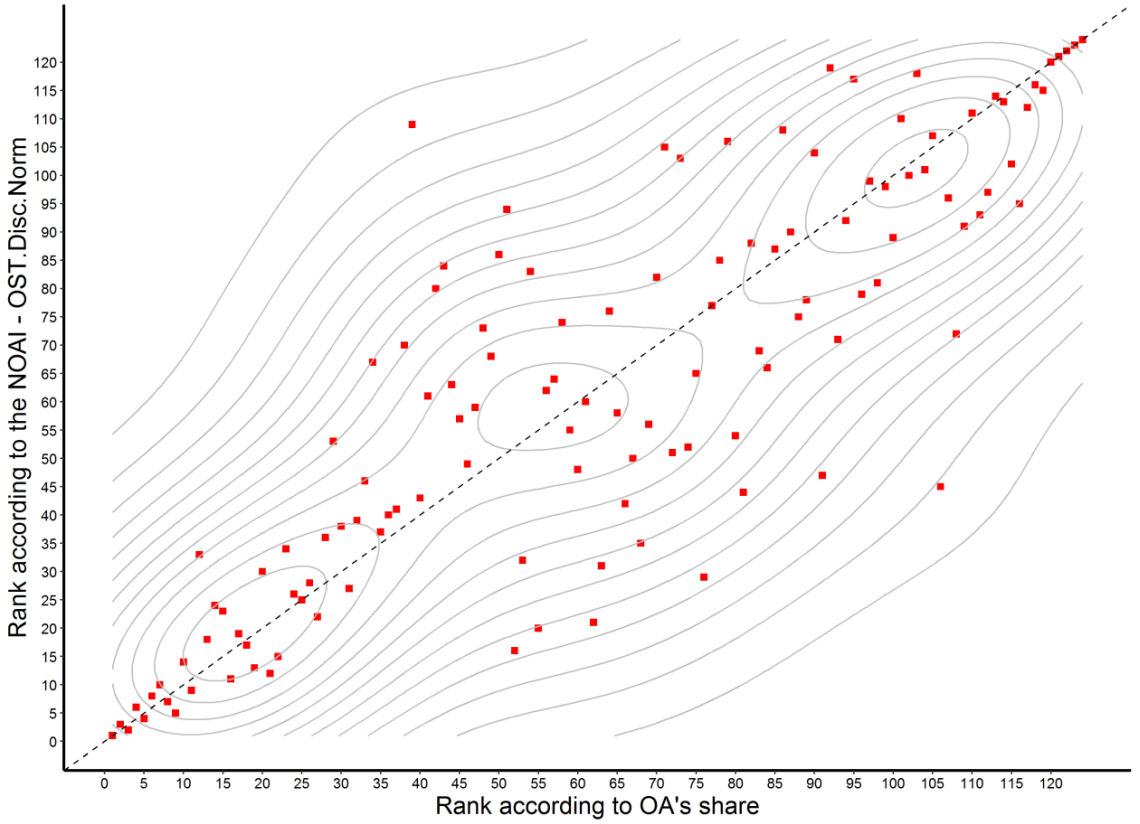


Figure 8 shows that when normalization is carried out by OST disciplines, rank changes are relatively less important. The correlation coefficient of the two ranks (OA share and NOAI) is higher (0.85).

Conclusion

Through this paper, we have shown that OA publishing practices vary by discipline. The rate of "openness" is relatively high in the disciplines of life sciences such as basic biology or medicine. The rate is much lower in engineering or computer science.

When it comes to making institutional (or country) comparisons it is imperative to take into account their disciplinary specificities and their specialization. The normalized indicator "NOAI" proposed in this paper consists in reporting at first the share of OA publications in a given discipline for an actor (institution, country, etc.) on the same part at the world level. In a second step, calculate a weighted average of normalized OA shares by discipline. This allows having an overall indicator of OA corrected for disciplinary differences in terms of openness.

Two levels of aggregation are used for normalization. A more general level of aggregation comprising 11 disciplines (OST nomenclature). And a fine aggregation level represented by the 255 disciplinary categories of the WoS database. The results indicate that the normalization obtained using the second level allows for better accuracy.

Limitations

1. The results are for the year 2016 and may be relatively volatile, especially for smaller institutions (use instead periods).
2. The shares of OA do not necessarily reflect the true practices of the disciplines, but rather give an image of the base WoS. Normalization allows making a correction (that remains insufficient).
3. There needs to be more discussion about the type of account to use. Geographical fractioning does not conceptually make much sense for this indicator (but allows making sums).

ACKNOWLEDGEMENTS

The author would like to thank Yves Gingras, Vincent Larivière and Lesya Baudoin for their comments on an earlier version and for their guidance in improving the manuscript.

References

Anderson. (2017a). When the wolf finally arrives: big deal cancelations in North American Libraries. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2017/05/01/wolf-finally-arrives-big-deal-cancelations-north-american-libraries/ (accessed 27 September 2019)

Anderson. (2017b). The forbidden forecast: thinking about open access and library subscriptions. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2017/02/21/forbidden-forecast-thinking-open-access-library-subscriptions/ (accessed 27 September 2019)

Antelman Kristin. (2004). "Do Open-Access Articles Have a Greater Research Impact?" College and Research, Libraries 65: 372–382. https://doi.org/10.5860/crl.65.5.372.

Antelman K. (2017). Leveraging the growth of open access in library collection decision making. In: Proceeding from ACRL 2017: at the helm: leading transformation.

Archambault É, Amyot D, Deschamps P, Nicol AF, Provencher F, Rebout L, Roberge G. (2014). Proportion of open access papers published in peer-reviewed journals at the European and world levels–1996–2013. European Commission

Borrego, Á. (2016). Measuring compliance with a Spanish Government open access mandate. Journal of the Association for Information Science and Technology, 67(4), 757–764. https://doi.org/10.1002/asi.23422

Björk, B.-C. (2017). Gold, green, and black open access. Learned Publishing, 30(2), 173–175. https://doi.org/10.1002/leap.1096

Björk B.-C (2016a). Hybrid open access—a longitudinal study. Journal of Informetrics 10(4):919-932.

Björk, B.-C. (2016b). The open access movement at a crossroad: Are the big publishers and academic social media taking over? Learned Publishing, 29(2), 131–134. https://doi.org/10.1002/leap.1021

Björk B.-C, Laakso M, Welling P, Paetau P. (2014). Anatomy of green open access. Journal of the Association for Information Science and Technology 65(2):237-250

Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. PLoS ONE, 5(6), e11273. https://doi.org/10.1371/journal.pone.0011273

Bolick, J. (2017). Exploiting Elsevier's Creative Commons License Requirement to Subvert Embargo. Poster Session Presented at the Kraemer Copyright Conference. Retrieved from http://hdl.handle.net/1808/24107

CWTS Leiden Ranking (2019). https://www.leidenranking.com/ranking/2019/list

Eysenbach G. (2006). Citation advantage of open access articles. PLoS Biol 4(5): e157. DOI: 10.1371/journal.pbio.0040157.

Gargouri Y, Larivière V, Gingras Y, Carr L, Harnad S. (2012). Green and gold open access percentages and growth, by discipline.

Harnad S. and Brody T. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. D-Lib Magazine, 10 (6).

Laakso M, Björk BC. (2012). Anatomy of open access publishing: a study of longitudinal development and internal structure. BMC Medicine 10 Article 124.

Laakso M, Björk B. 2013. Delayed open access: an overlooked high-impact category of openly available scientific literature. Journal of the American Society for Information Science and Technology 64(7):1323-1329.

Martín-Martín A., Costas R., Van Leeuwen T., López-Cózar E-D. (2018). Evidence of Open Access of scientific publications in Google Scholar: a largescale analysis, Journal of Informetrics, vol. 12, no. 3, pp. 819-841, 2018. https://doi.org/10.1016/j.joi.2018.06.012

Piwowar H., Priem J., Larivière V., Alperin JP., Matthias L., Norlander B., Farley A., West J., Haustein S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. PeerJ 6, e4375. (doi:10.7717/peerj.4375)

Schimmer R., Geschuhn K. K., & Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. doi:10.17617/1.

Schiermeier Q, Mega ER. 2017. Scientists in Germany, Peru and Taiwan to lose access to Elsevier journals. Nature News 541(7635):13

Smith E, Haustein S, Mongeon P, Fei S, Ridde V, Larivière V. (2017). Knowledge sharing in global health research; the impact, uptake and cost of open access to scholarly literature. BMC Health Research Policy and System In Press.

Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CH. 2016. The academic, economic and societal impacts of Open Access: an evidence-based review (version 3; referees: 3 approved, 2 approved with reservations) F1000 Research 5 Article 632

Universitat Konstanz. (2014). Teurer als die Wissenschaft erlaubt.

Université de Montréal. (2017). UdeM Libraries cancel Big Deal subscription to 2231 periodical titles published by Taylor & Francis Group.

Walker TJ., Soichi transl. T. (1998). Free internet access to traditional journals. Journal of Information Processing and Management 41(9):678-694.

Appendices

1. OST's disciplines

Major disciplines	Abbreviations
Applied biology - Ecology	App. Bio. - Eco.
Fundamental biology	Fund. bio.
Chemistry	Chemistry
Computer science	Comp. Sc.
Mathematics	Maths
Physics	Physics
Medical research	Medical R.
Engineering	Engineering
Earth sciences –	Earth sc., Astro.
Humanities	Humanities
Social sciences	Soc. Sc.

3. ERC Panels

ID	Wording
SH1	Individuals, Markets and Organizations
SH2	Institutions, Values, Environment and Space
SH3	The Social World, Diversity, Population
SH4	The Human Mind and Its Complexity
SH5	Cultures and Cultural Production
SH6	The Study of the Human Past
PE1	Mathematics
PE2	Fundamental Constituents of Matter
PE3	Condensed Matter Physics
PE4	Physical and Analytical Chemical Sciences
PE5	Synthetic Chemistry and Materials
PE6	Computer Science and Informatics
PE7	Systems and Communication Engineering
PE8	Products and Processes Engineering
PE9	Universe Sciences
PE10	Earth System Science
LS1	Molecular Biology, Biochemistry, Structural Biology and Molecular Biophysics
LS2	Genetics, 'Omics', Bioinformatics and Systems Biology
LS3	Cellular and Developmental Biology
LS4	Physiology, Pathophysiology and Endocrinology
LS5	Neuroscience and Neural Disorders
LS6	Immunity and Infection
LS7	Applied Medical Technologies, Diagnostics, Therapies and Public Health
LS8	Ecology, Evolution and Environmental Biology
LS9	Applied Life Sciences, Biotechnology, and Molecular and Biosystems Engineering