State of Open Access penetration in universities worldwide

Nicolas Robinson-Garcia[1], Rodrigo Costas[2,3] and Thed N. van Leeuwen[2]

[1] Delft Institute of Applied Mathematics (DIAM), TU Delft, Delft, Netherlands
[2] Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, Netherlands
[3] Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

Corresponding Author:
Nicolas Robinson-Garcia[1]
Building 28, Mourik Broekmanweg 6, 2628 XE Delft, Netherlands
Email address: elrobinster@gmail.com

Abstract

The implementation of policies promoting the adoption of an Open Science culture must be accompanied by indicators that allow monitoring the penetration of such policies and their potential effects on research publishing and sharing practices. This study presents indicators of Open Access (OA) penetration at the institutional level for universities worldwide. By combining data from Web of Science, Unpaywall and the Leiden Ranking disambiguation of institutions, we track OA coverage of universities' output for 963 institutions. This paper presents the methodological challenges, conceptual discrepancies and limitations and discusses further steps needed to move forward the discussion on fostering Open Access and Open Science practices and policies.

Introduction

The implementation of policies promoting the adoption of an Open Science culture must be accompanied by indicators that allow monitoring the penetration of such policies and their potential effects on research publishing and sharing practices. In this paper we present Open Access (OA) indicators for universities worldwide. We analyse the presence of OA by type of access, field differences and comparisons with scientific impact and international collaboration. We explore discrepancies between the operationalization of OA indicators and the conceptualization of OA.

The notion of Open Science goes back to the sixteenth Century (David, 2008), but it has recently gained relevance as the EU introduced it as a pivotal stone in their research programmes (Moedas, 2015). Within the different directives set up to achieve it, OA has become one of the first milestones. Initiatives such Plan S (Else, 2018a,b) or the European Commission's Open

Science Monitor[1] exemplify such efforts and the prioritization of OA for these agencies. The latter being the tool the European Commission is using to monitor its penetration. However, more granular levels of analysis are needed to better understand how OA is expanding, which OA models are being implemented and what are the potential side-effects of such models. Universities have been supporting OA for many years now. The most common has been by building and maintaining institutional repositories, and introducing mandates that oblige their researchers to deposit their publications (Harnad, 2007; Harnad et al., 2008). There is also evidence of institutions promoting OA publications by sponsoring the costs derived from the article processing charges (APC) of open journals (Gorraiz & Wieland, 2009; Gorraiz, Wieland & Gumpenberger, 2012). In most cases, institutions are faced with the challenge of determining the success of such initiatives and monitoring the compliance of their researchers with international and national OA mandates. Initiatives such as the ranking of OA repositories (Aguillo et al., 2010) offer partial information on the share of OA available at the institutional level. Although valuable, is still insufficient, as institutional repositories may not be the main vehicle used by researchers to make their outputs openly accessible (Arlitsch & Grant, 2018), and not all the researchers at the same university comply with their institutional mandates in the same manner.

Until five years ago, there were no more than estimates as to the amount of OA publications. However, the development of platforms like CrossRef, DOAJ or even Google Scholar, along with computational advancements on web scrapping, have led to a plethora of large-scale analyses to empirically identify OA literature (Archambault et al., 2014; van Leeuwen, Tatum & Wouters, 2018; Piwowar et al., 2018; Martín-Martín et al., 2018b). Overall, these studies report that around half of the scientific literature is freely available but point towards the increasing availability of publications which do not adhere strictly to what is considered OA. The game changer in this respect, has been Unpaywall (Piwowar et al., 2018), a product developed by the non-profit Our Research[2], which tracks OA versions of published research, recently becoming the most standard mechanism to identify OA

In this paper, we present a first attempt at analyzing OA at the institutional level. The purpose is twofold. First, to inform how OA is being achieved in different institutions and countries, describe national trends, and pathways by which OA is being expanded. Second, we deepen into green and gold OA types to analyze both empirical and conceptual discrepancies on how these two types of OA are understood. The results of this study have been recently incorporated to the 2019 edition of the Leiden Ranking released in May 2019 (van Leeuwen, Costas & Robinson-Garcia, 2019) and a first version was presented at the ISSI 2019 Conference (Robinson-Garcia, Costas & van Leeuwen, 2019).

[1] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en

[2] https://ourresearch.org/

Materials & Methods

In this paper we use different sets of sources and combine different methods to determine OA. Publication data is retrieved from the CWTS in-house version of the Web of Science. Unlike in the Leiden Ranking (restricted to article and reviews), here we report indicators for letters, articles and reviews indexed in the Science Citation Index Expanded, Social Sciences Citation Index and Arts & Humanities Citation Index for the 2014-2017 period. We link publications to the 963 universities identified in the Leiden Ranking database via their disambiguated list of institutional names, also hosted at CWTS (Waltman et al., 2012). Publications are assigned to five fields of science, following the methodology employed in the Leiden Ranking[3]. These fields are: Biomedical and Health Sciences, Life and Earth Sciences, Mathematics and Computer Science, Physical Sciences & Engineering, and Social Sciences and Humanities.

For each publication, we identify if they are openly accessible and the type of Open Access by querying the Unpaywall information. Unpaywall relies on Digital Object Identifiers (DOI), which means that we will only include records which have a DOI assigned to them. Furthermore, the Unpaywall API does not label types of OA but records different pieces of evidence of OA availability of each publication. More information on the Unpaywall approach to OA is available at their User Guide offered for researchers (http://unpaywall.org/data-format).

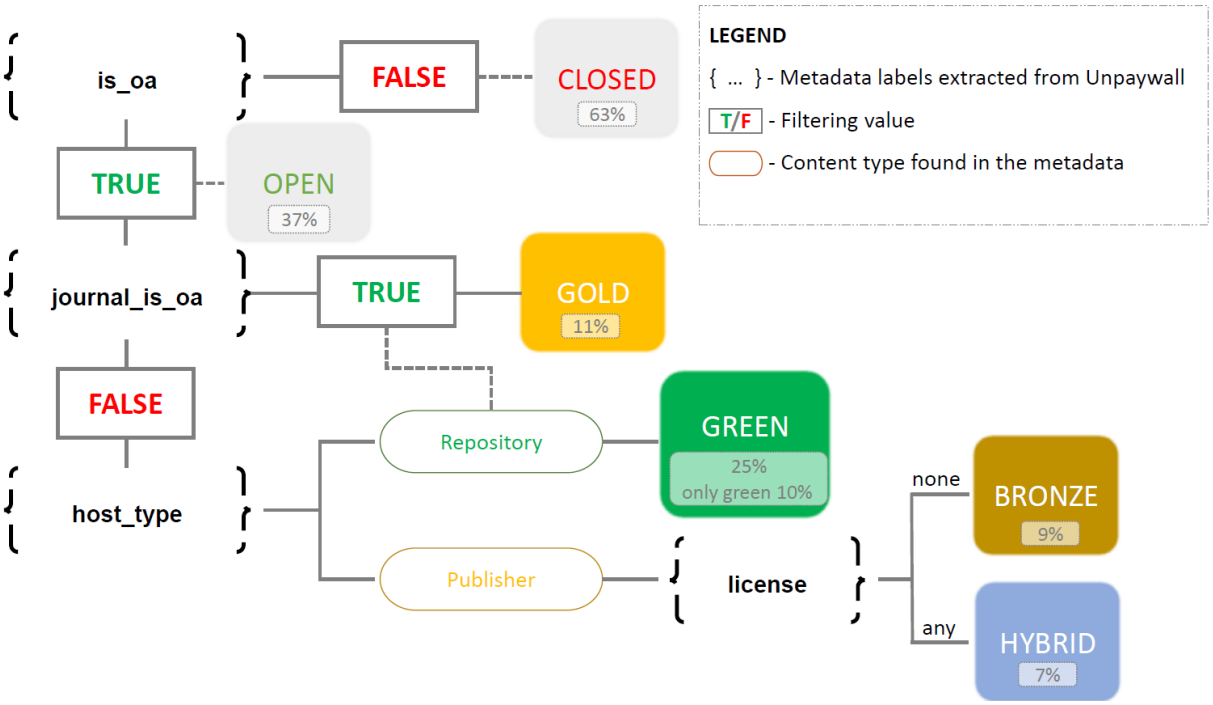

Figure 1 Workflow followed to identify OA types based on Unpaywall data. Source: van Leeuwen, Costas and Robinson-Garcia 2019

[3] A detailed description the assignment of publications to fields is provided here https://www.leidenranking.com/information/fields

Four types of OA are considered. These four types of OA are defined as follows:

- **Green OA.** Self-archived versions of a manuscript. Here the responsibility lies on the authors of the publication, or institutional colleagues such as central library staff members, who oversee depositing the document in a repository. This version of the document may not correspond with the final version of the publisher.
- **Gold OA.** This refers to journals which publish all their manuscripts in OA regardless of the business model they follow (e.g., publicly sponsored, author pays).
- **Hybrid OA**. Toll access (non-OA) journals make specific publications openly accessible usually after the author pays a fee to account for potential losses derived from subscription fees.
- **Bronze OA.** Again, toll access journals are the ones offering the publication freely available, however this OA is not subjected to copyright conditions set to be defined as OA (i.e., they do not ensure perpetual free access).

The labelling of OA types is described in Figure 1 and already highlights some of the difficulties raised when trying to define what is actually OA (Torres-Salinas, Robinson-Garcia & Moed, 2019). The Unpaywall API provides for each publication record a set of different pieces of OA evidence. For each piece of evidence, we study all the metadata labels referring to the OA status of the publication. Thus, when one piece of evidence suggests that a paper belongs to an OA journal (gold OA), this automatically overrides bronze or hybrid OA, since conceptually gold, bronze and hybrid are incompatible. The only exception made with green OA, which could overlap with any of the other three types.

Overall, a total of 4,621,721 distinct publications records are examined, out of which 1,881,192 records were identified as OA (40.7%). Figure 2 shows how these OA publications are distributed by type. 77% of all OA publications were green OA, followed by gold OA (33%), bronze OA (20%) and hybrid OA (16%). However, there is a substantial overlap between each of these latter OA types and green OA. 81% of all gold OA publications are also in green OA, for hybrid the share which is also green is 63%, and of hybrid are 45% for bronze OA.

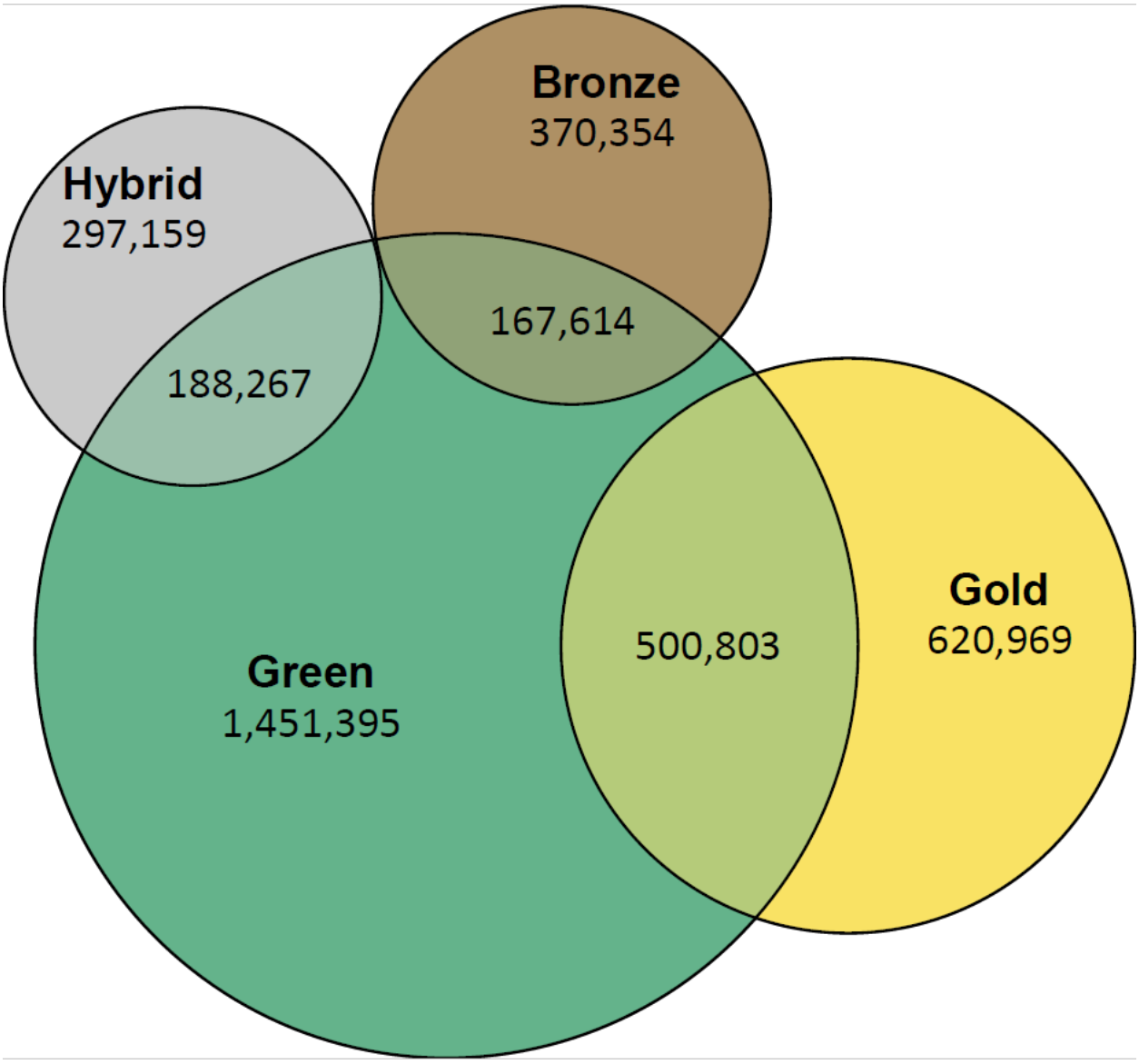

Figure 2 Total number of documents in open access by type and overlap.

The results are reported at different levels. First analyses investigate the share of OA on the overall output of each university, differences by country, continent and field. We then look specifically into the contents of what is regarded as green OA. For this we explore where are green OA publications stored by querying the Unpaywall API for the specific path where the OA document was retrieved from. Finally, we different national models of gold OA publishing. We characterize gold OA publishing based on three variables: share of papers published in national journals, share of papers published in English language and share of papers published in journals following including Article Processing Charges (APC). Language of documents and journal's country are identified using data from Web of Science. In the case of the latter, we identify the country of the journal by querying the field Publisher Address (PA).

In the case of APCs, we queried the Directory of Open Access Journals. Here we must note that this is not a comprehensive list of OA journals. Unpaywall identifies a larger number of gold OA journals (n= 11,601) than DOAJ (n= 11,365), and for which we have no information on APCs. Therefore, the numbers on gold OA journals with/out APCs provided represent a lower bound of all the gold OA journals for which APC information is available via DOAJ. A total of 768 APC journals were identified. After some inspection, we found some inconsistencies in the way APC is defined according to DOAJ. That is, not in all cases, APC refers to an author pays model, but in some cases, journals offer an optional subscription fee for those interested on accessing to printed versions of the journal. This is the case for many journals stored in the SciELO platform which are free of costs for both readers and authors but give the option to pay a subscription fee for printed versions of the journal.

Results

General overview

In Figure 3 we consider the proportion of OA publications by countries. Only countries with at least 10 universities listed in the Leiden Ranking are shown. The median share of publications openly available of universities worldwide is 43%. British universities have by far the largest share of OA publications (median=74%), followed by Sweden (median=56%) and Austria (median=54%). Except for the United States (median=51%) and Brazil (median=47%), all countries above world median are European. Asian countries, as well as Canada and Australia show OA shares below the world median.

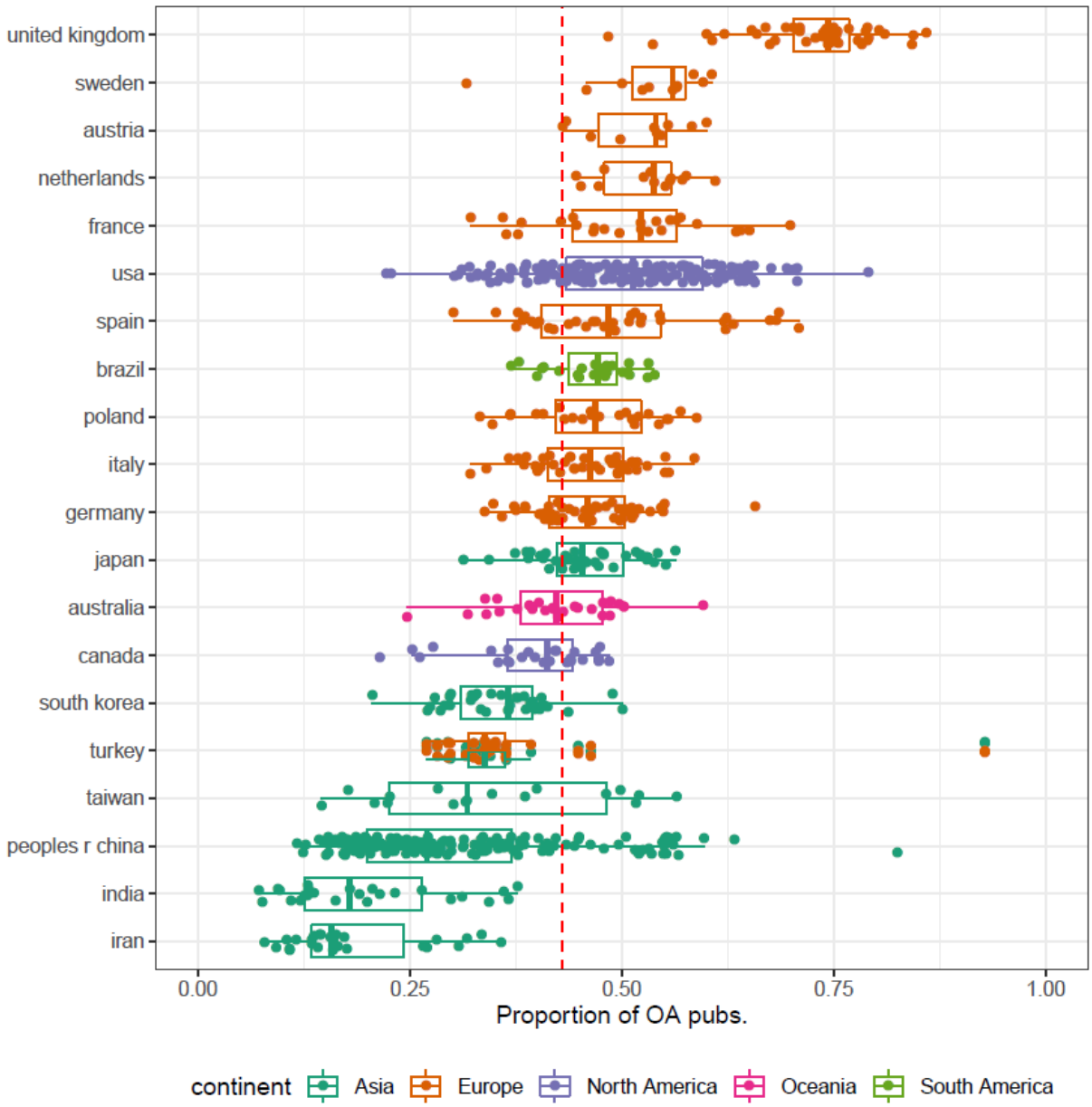

Figure 3 Proportion of OA publications of the set of universities analysed by countries. Only countries with at least 10 universities included are shown. Countries are ordered based on the median value of the share of OA publications of their universities. The red dashed line indicates the world median value. Turkey is assigned to both, Europe and Asia.

We disaggregate by type of OA in Figure 4. Most OA publications are openly accessible via the green route, and hence the similarity between Figure 3 and Figure 4A. In the case of gold OA (Figure 4B) a very different image is seen. Brazilian universities outstand with a median of 30% publications in Gold OA. Sweden is placed in second, along with Taiwan (median=18% for both countries). Universities from United Kingdom (median=17%), Austria (median=15%) and Netherlands (median=13%) correspondingly, show the highest share of hybrid OA publications.

While for bronze OA, it is universities from Japan (median=15%), Turkey (median=13%) and Netherlands (median=12%) the ones outstanding.

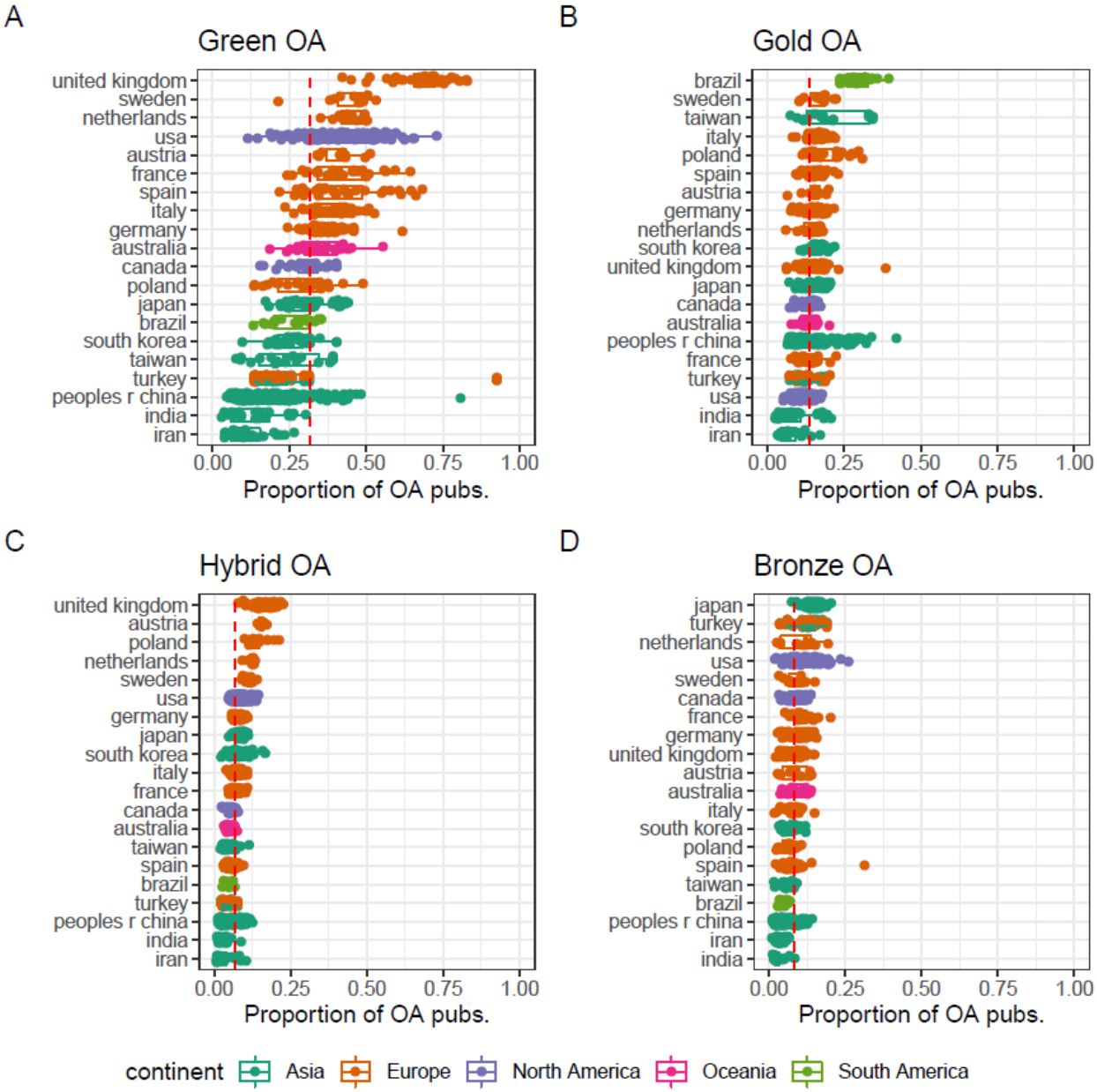

Figure 4 Proportion of OA publications of the set of universities analysed by countries for each type of OA. Only countries with at least 10 universities included are shown. Countries are ordered based on the median value of the share of OA publications of their universities. The red dashed line indicates the world median value. Turkey is assigned to both, Europe and Asia.

Figure 5 shows the predominance of each OA type by field and at the university level (each point represents the share of a university in each field and type of OA grouping). The average share of

OA publications is 42.8%. The largest median is found in the Biomedical and Health Sciences (49.1%), while Social sciences and Humanities exhibit the lowest shares of OA (36.5%). Green OA is the most predominant form of open access regardless of the field (median of 33.2% in the 'All sciences' group). Again, the largest average is found in Biomedical and Health Sciences (39.0%) and the lowest in Social sciences and Humanities (28.0%). Overall, universities publish on average 14.7% of their publications in OA journals. For Biomedical & Health Sciences the average increases up to 19.3%, while in Mathematics & Computer Science it drops to 9.0%. In the case of Hybrid OA, an average of 7.1% of papers in universities are published under this modality. This figure increases in the case of Physical Sciences and Engineering to 7.9%, while in Social sciences & Humanities it represents an average of 4.6% of the output. Bronze OA, although it not strictly OA as it does not ensure sustainable access, is more common on average than Hybrid OA, with an overall share of 8.5% which goes up to 11.1% for Biomedical & Health Sciences, but with a presence on average of 3.7% in Mathematics & Computer Science.

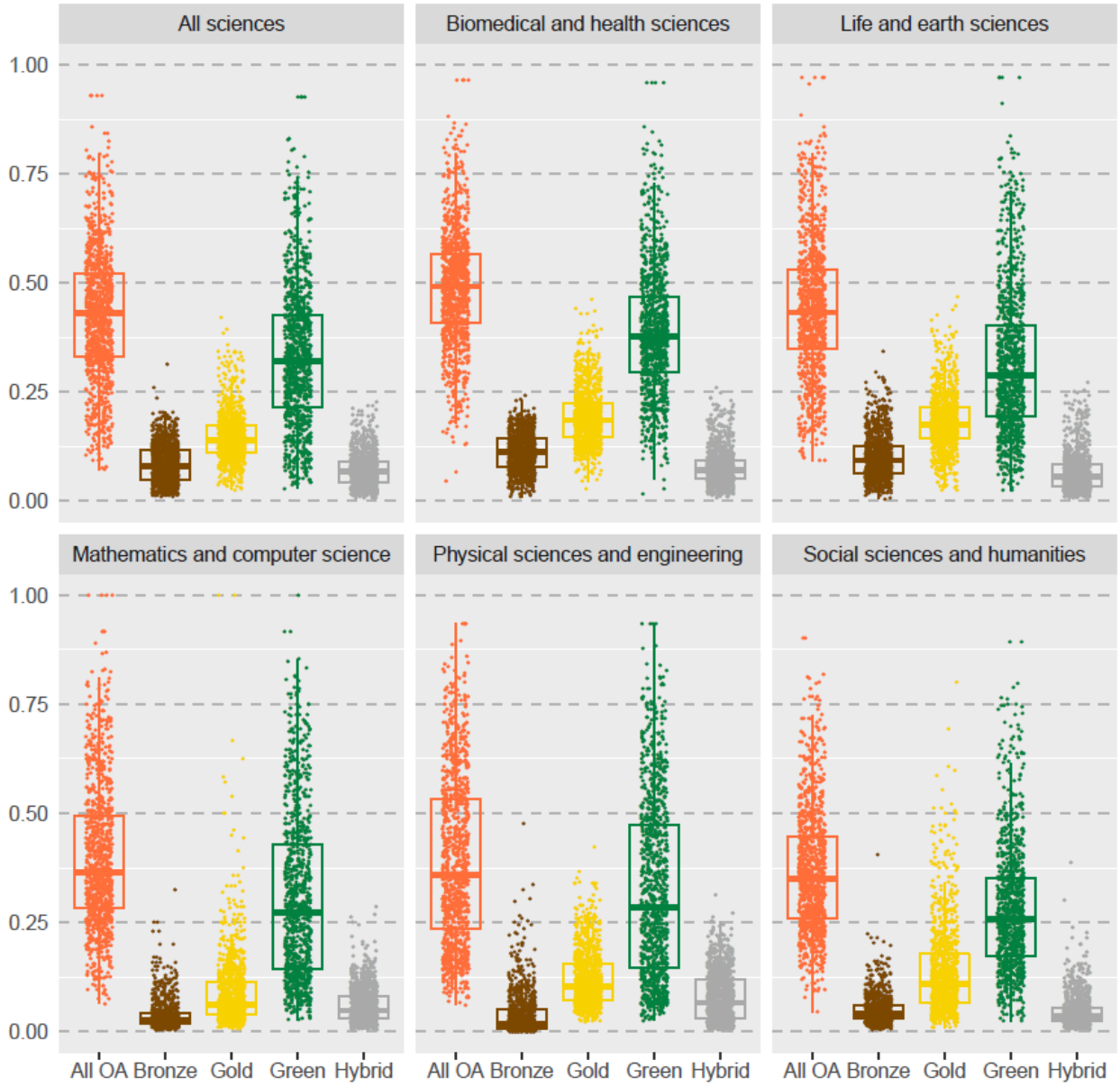

Figure 5 Proportion of OA publications of universities for each type of OA and for all OA types by field and B) region for universities worldwide.

We also note large differences by geographical region (Figure 6). Europe (50.1%) and North America (49.1%) are the continents with the universities sharing the largest proportions of their output in OA. In the other extreme we find Asia (32.5%) and Africa (39.1%). In the former two continents, green OA is by large the most common OA type (41.1% in Europe and 40.6% in North America) with gold OA lagging behind by far as the second option (15.4% and 12.0% respectively). In South America, median shares of green (29.2%) and gold OA (27.0%) by university are practically identical. Shares of hybrid and bronze OA are on median below 10% for all continents except for bronze OA in North America (11.2%).

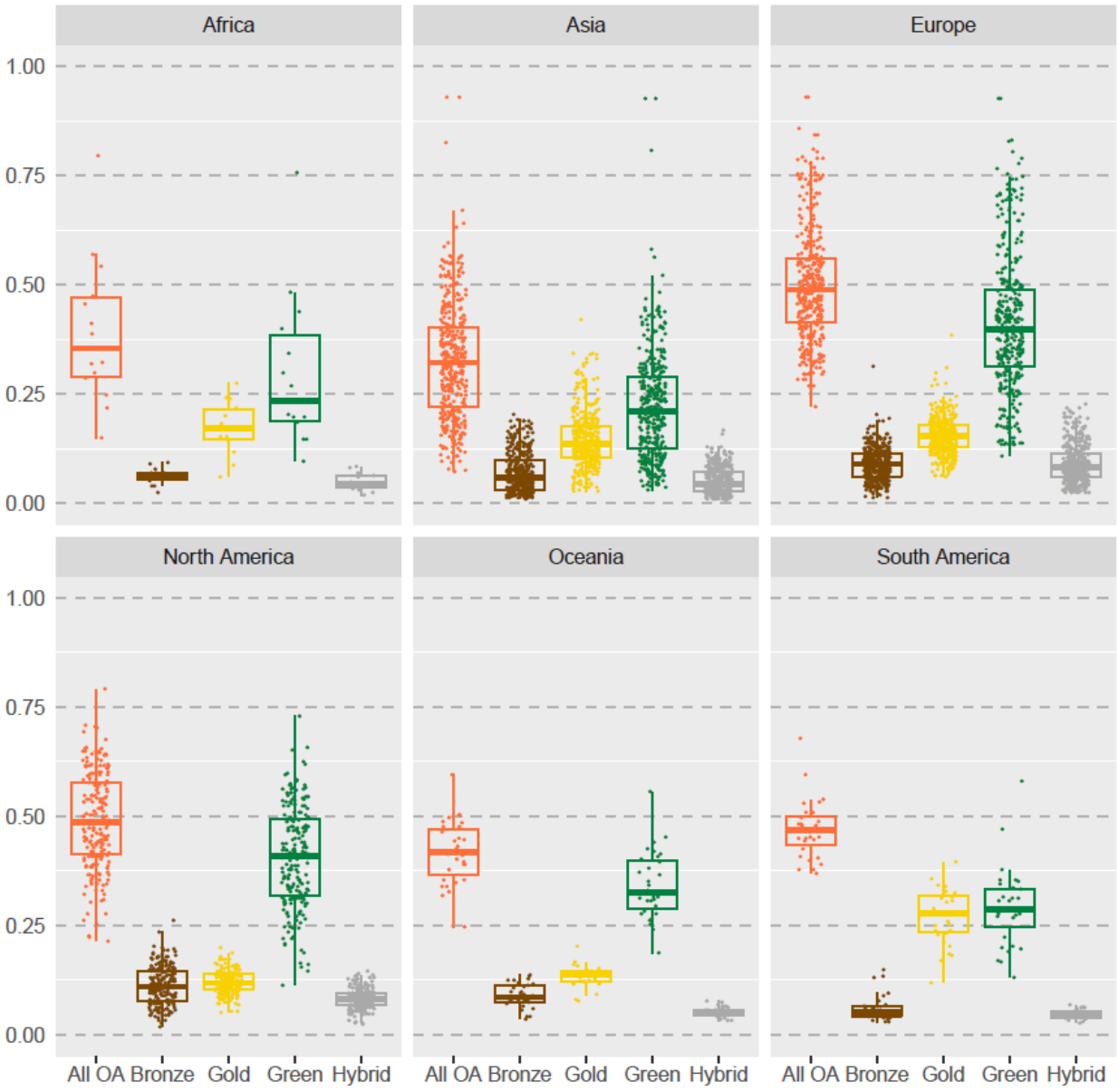

Figure 6 Proportion of OA publications for each type of OA and for all OA types by region for universities worldwide.

University profiling

It is remarkable that differences between and within universities can be quite significant. In Figure 7 we take a closer look into the disciplinary profile of a set of universities based on the type and proportion of OA output by field. To illustrate the OA institutional profiling of universities, we use radar charts and select in each row the three universities with the largest output (considering their full counting) in North America, Europe, Africa, South America and Asia, respectively. In the first row, we observe the three largest universities in North America,

two from the United States and one from Canada. The two US universities have above half of their output in green OA, with Social sciences and humanities, just below the 50% threshold. In the case of the University of Toronto, the shares are much lower, ranging between 39% green OA in Biomedical and Health Sciences and 3% bronze OA in Mathematics and Computer Science. The three largest universities in Europe are all from the United Kingdom. Again, green OA is clearly the most common OA option in all fields within these universities, showing more homogeneity across the three institutional profiles. However, Social Sciences and Humanities tend to have lower shares for the universities of Cambridge and Oxford than for University College London.

Figure 7 An example of OA disciplinary profiles for top 3 universities with the largest output for North America, Europe, Africa, South America and Asia.

Regarding Africa (third row), two of the three universities showcased are South African, while the third one is Egyptian. In the case of Cairo University, no OA type in any field reaches a quarter of the total output of the university. In the other two cases, the profiles are quite similar, with the University of Cape Town exhibiting higher shares of green OA than the University of Witwatersrand. For South America, three Brazilian universities outstand as the largest ones; Universidade de São Paulo, Universidade Estadual Paulista and Universidade Estadual de Campinas. The gold OA preponderance previously observed at an aggregate level both for the continent and Brazil, is also noted at the institutional level in all three universities. However, we do observe that such preponderance is coming mainly from the Life and Earth Sciences and the Social Sciences and Humanities. Finally, for Asia (last row), we profile three Chinese universities for which green and gold OA shares go hand in hand in all three cases, with the exception of the field of Biomedical and Health Sciences, where green OA reaches higher shares of the total output.

Green Open Access and self-archiving

We will now delve into green OA, to better comprehend the indicators shown displayed on this typology. Green OA was originally defined as self-archiving of preprint or post print versions of published manuscripts. That means that green OA is achieved as the result of a proactive attitude of the authors or an institutional colleague, like librarians, towards OA. In their seminal paper, Harnad et al. (2004) go beyond such definition, and indicate that "the self-archiving method with the greatest potential to provide OA is self-archiving in the author's own university's OAI-compliant Eprint Archives" (p. 312). Hence, one could expect to see in the green OA indicators, shares of institutional self-archiving of a university's output. However, a closer look into what is considered as green following the identification procedure used based on Unpaywall data, shows that this is not the case for two reasons.

First, the assignment of OA output to each university is given based the affiliation of authors and not the contents of institutional repositories. This means that universities with large proportions of their output in green OA may not be succeeding on storing their output in their institutional repositories themselves. Table 1 shows the top 20 universities with the largest shares of their output in green OA. Along with the total number of publications and green OA publications, we provide a threshold of the share of publications which are stored in their own institutional repository. We identify the lower band of the threshold by individually querying the URL string of each university's repository. The upper band results from also including URL string containing hdl.handle.net, which is the URL used when linking through the HANDLE identifier, a similar identifier to DOIs but assigned by repositories. Some universities do have most of their output accessible thanks to their own institutional repositories. For instance, 98% of green OA publications from Bilkent University are stored in their own repository.

University	Country	Pubs	Green pubs	Pubs in Repository*
Bilkent Univ	Turkey	2,008	1,858	1,815 – 1,815
City Univ London	United Kingdom	2,569	2,131	1,881 – 1,888
Durham Univ	United Kingdom	7,452	6,159	5,227 – 5,239
Hong Kong Polytech Univ	China	9,816	7,925	406 – 7,626
London Sch Hyg	United Kingdom	7,237	5,817	4,434 – 4,459
Univ Strathclyde	United Kingdom	4,847	3,830	3,403 – 3,409
Univ St Andrews	United Kingdom	5,780	4,497	3,562 – 3,584
Loughborough Univ	United Kingdom	4,274	3,271	2,810 – 2,820
Univ Pretoria	South Africa	6,432	4,873	4,564 – 4,567
Univ Leeds	United Kingdom	11,948	8,994	7,379 – 7,401
Univ Glasgow	United Kingdom	12,024	8,975	6,993 – 7,029
Univ Bath	United Kingdom	5,142	3,808	2,600 – 2,623
Univ Edinburgh	United Kingdom	18,139	13,415	7,401 – 7,810
Caltech	United States	13,481	9,834	6,804 – 6,821
Univ Bristol	United Kingdom	14,297	10,418	6,493 – 6,545
Univ Reading	United Kingdom	4,720	3,408	2,888 – 2,894
London Sch Econ	United Kingdom	3,525	2,534	2,013 – 2,023
Univ Coll London	United Kingdom	35,352	25,366	15,770 – 15,878
Univ Sussex	United Kingdom	5,510	3,931	2715 – 2,738
Univ Warwick	United Kingdom	10,706	7,644	4538 – 4,580

Table 1 Table 1. Top 20 universities with the highest share of their output available through green OA.

****The interval refers to: lower bound when querying only for the institutional repository's URL string, and upper bound when querying for the institutional repository's URL string or hdl.handle.net. When searching for the _hdl.handle.net_ string, the share increases to 73.8% of the total output.***

Second, low coverages of green OA output in institutional repositories can be due to inter-institutional collaboration (i.e. collaboration with other institutional partners that apply more systematic archiving policies) or self-archiving in thematic (e.g., ArXiv) or supranational repositories (e.g., Zenodo). However, there is a second phenomenon which drifts further away the original definition of green OA from the actual numbers that are reported based on the general labelling obtained via Unpaywall. That is, the effect of repositories which store OA documents without authors' intervention. Previously, we referred to this as different perspectives of green OA based on "the degree of engagement" of the authors (van Leeuwen, Costas & Robinson-Garcia, 2019). We distinguish between two perspectives: 1) self-archiving, defined as the deliberate action of an author or librarian to archive publications in a repository, and 2) general archiving, where the archival function is still taking place, but without the explicit intervention of the author or librarian. So far, we have identified one macro repository following this general archiving perspective; PubMed Central (https://www.ncbi.nlm.nih.gov/pmc/). This

source alone represents 60.8% of the green OA literature identified. However, some of its contents are retrieved from elsewhere, including OA journals such as Plos ONE. 86.5% of the 881,834 documents in PMC are simultaneously also gold, bronze or hybrid OA. The remaining 13.5% is accessible via another repository as well as PMC. As it is indeed a repository, in this study it is considered as a green OA source, but the effect of such decision in OA shares at the institutional level is highly significant. Figure 8 shows the effect of PMC on the shares of green OA. 49 universities are shown, these are those for which green OA deposited in PMC represents 95% or more of their total number of green OA publications. While self-archived and PMC publications can overlap (as more than one instance of OA evidence can be found per publication), in some cases the difference between defining PMC publications as green or not can derive on up to more than 10,000 publications, as in the case of University of Texas, Houston.

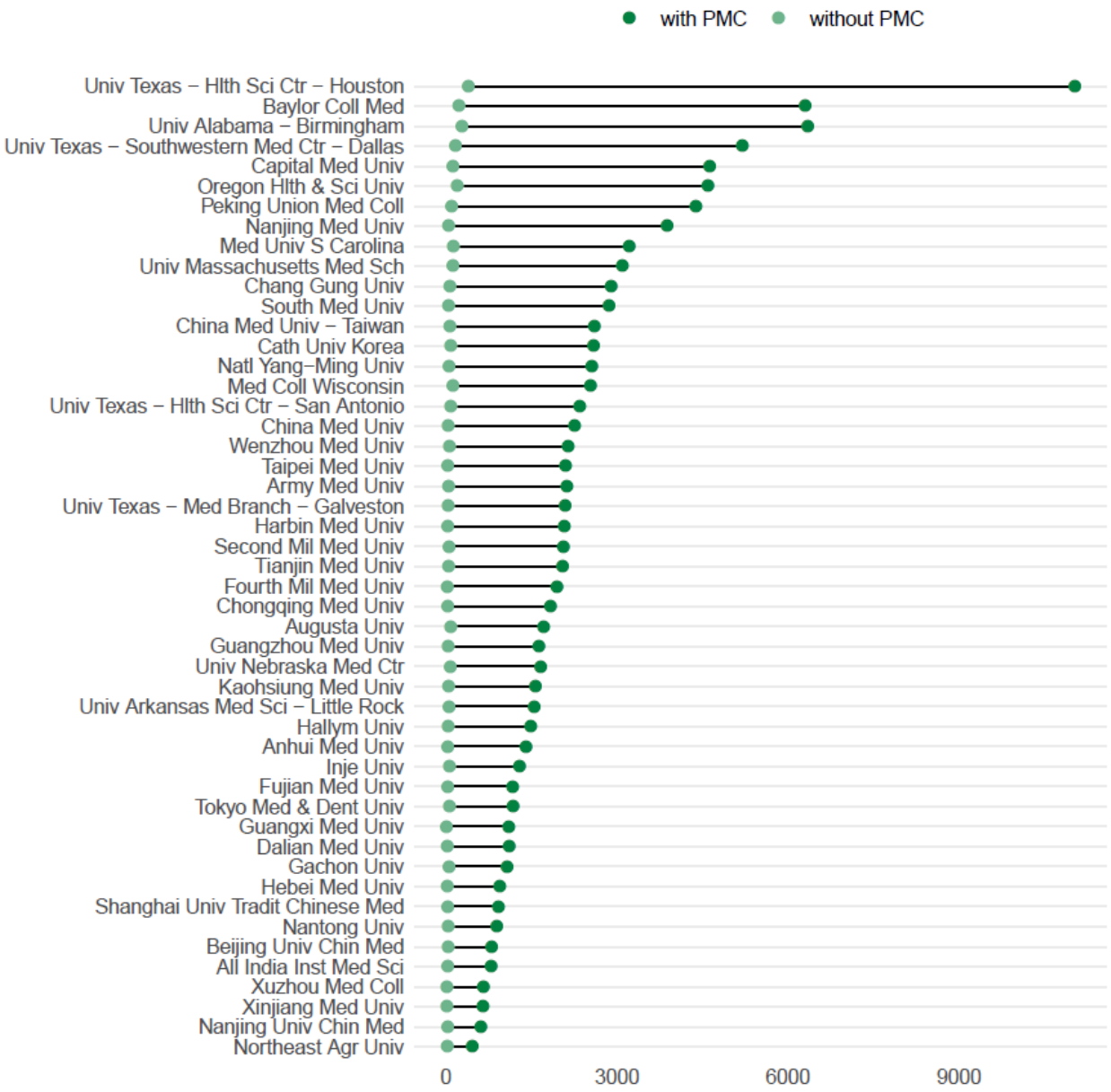

Figure 8 Difference on number of green OA publications with and without PMC. Only where PMC represents 0.95 or more of the share of green OA is shown

Country	Green OA	PMC	PMC only	% Gold	% Bronze	% Hybrid
Taiwan	18,841	14,748	12,337	825	77	111
South Korea	43,425	34,066	26,995	4,521	673	909
China	190,201	138,931	114,228	93,526	12,037	20,094
Thailand	5,166	3,987	2,578	2,434	444	497
Lebanon	819	620	386	383	68	69
Egypt	3,604	2,394	1,617	1,521	230	276
Japan	59,787	34,289	24,104	19,990	6,029	4,942

Country	Green OA	PMC	PMC only	% Gold	% Bronze	% Hybrid
Singapore	10,717	6,637	4,266	4,063	900	855
Malaysia	8,675	4,718	3,345	3,839	217	345
Poland	19,672	10,222	7,404	5,780	546	3,060
Pakistan	1,344	638	496	513	24	58
Austria	18,208	10,554	6,471	4,777	1,139	3,293
Canada	71,913	45,445	25,121	23,244	7,716	5,248
Iran	8,412	4,408	2,931	2,109	231	1,175
Brazil	35,134	18,901	11,707	14,395	1,398	1,152
India	10,475	4,923	3,414	3,305	424	407
USA	522,934	383,483	169,403	115,564	68,876	42,597
Israel	16,761	8,750	5,407	4,530	1,428	1,178
Mexico	6,133	2,758	1,924	1,982	284	188
Saudi Arabia	10,042	5,211	3,108	3,738	380	498

Table 2 Top 20 countries with the highest share of distinct green OA publications coming from PMC. Shares of gold, bronze and hybrid OA are based on the total number of PMC publications.

In Table 2 we aggregate the set of universities at the country level, to identify in which countries the inclusion of PMC as green OA affects the most their figures. The greatest effect is observed in Taiwan (65.5% of their total green OA), South Korea (62.2%) and China (60.1%). Furthermore, we observe that most of the documents coming from PMC are provided through another OA route, mostly gold, but also hybrid and bronze. This shows again the introduction of some degree of duplication of other OA types into green when including PMC and how the way we define and operationalize each of the OA types can affect the final numbers provided.

Gold Open Access models

As previously observed, Gold OA is the second largest type of OA of the four analysed here (Figure 2), but with some notable exceptions like the case of Brazil (Figure 4B). Torres-Salinas et al. (2019) highlight three models to characterize gold OA publishing from their analysis on Gold OA. The first one represents countries which publish in OA journals from big publishing firms and with a high Journal Impact Factor. Countries like United Kingdom, Germany or the Nordic countries fit into this model. A second model showcases countries publishing in national low Impact Factor OA journals, such as Brazil or India. The third model is a combination of the previous two, where they point out at countries like Poland or Spain. In Figure 9 we take a similar approach looking at three variables for Gold OA publishing: share of gold OA publications in APC journals, share of gold OA publications in English language and share of gold OA publications from national journals. We observe that patterns are quite stable for the three variables. Most countries publish up to 25% of their output in national OA journals. APCs

are paid for a range between 50% and 75% of their gold OA publications, and almost all of it is published in English language.

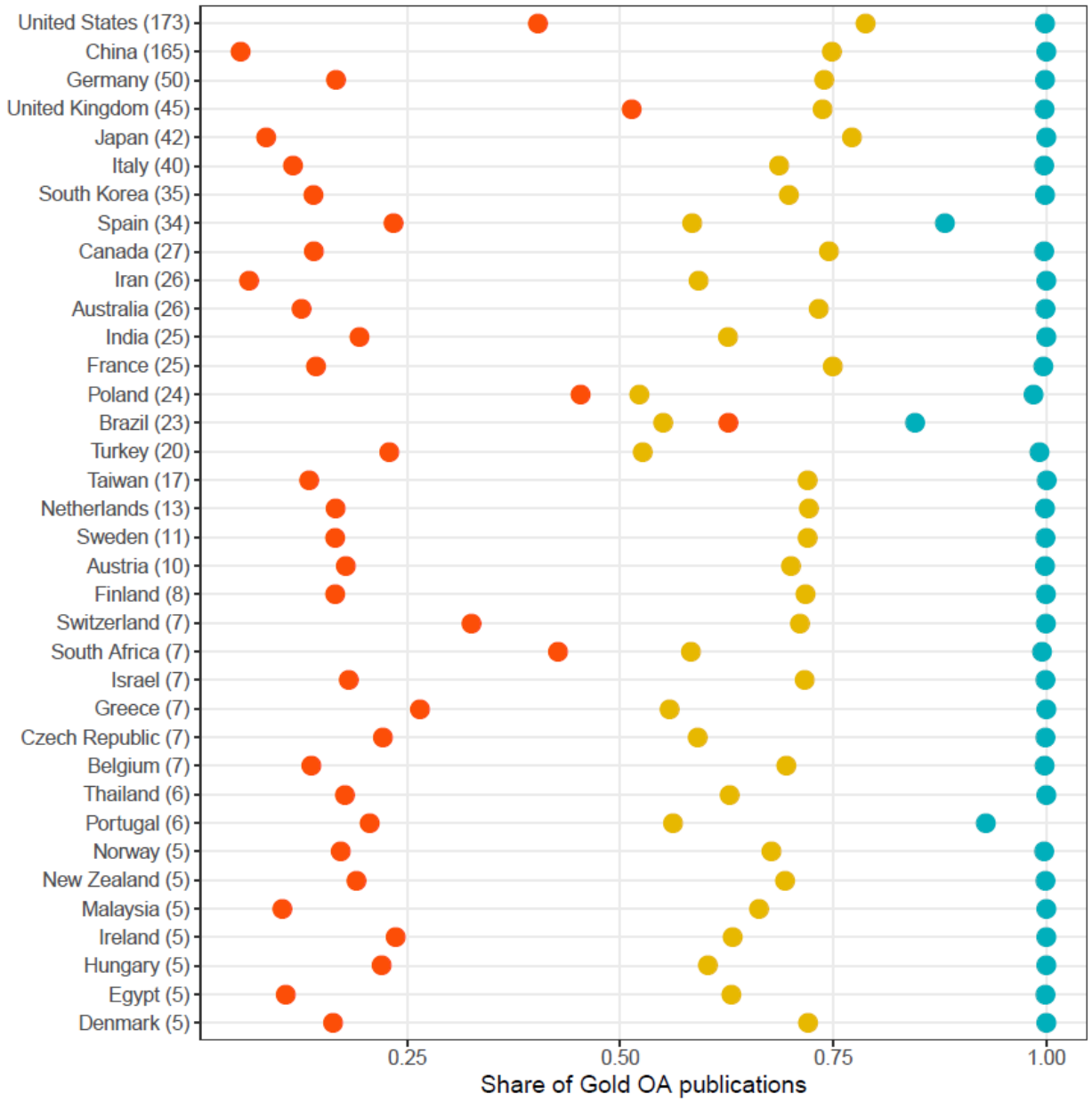

Figure 9 Share of Gold OA publications by country by type of publications. Orange: Publications in national OA journals; Yellow: Publications in APC OA journals; Blue: Publications in English language. APC data is extracted from the Directory of Open Access Journals (DOAJ). Only countries with at least 5 universities in the Leiden Ranking are shown.

This pattern is followed by most countries, but some differences can be observed. For instance, United States and United Kingdom represent countries with high level of APC publishing, high shares in national language and almost exclusively in English language. Switzerland also fits into

this pattern despite being a non-English speaking country. Another differing pattern is observed for countries like Spain or Portugal, where the share of English language publications is much lower although the share of national publications is still below 25% (23% and 21% respectively). In the case of Poland, although 98% of Gold OA publications are in English language, 45% come from national journals, with APC publications in the lower bound of the 50%-75% interval (52%). A similar pattern is followed by South Africa. Finally, we highlight the case of Brazil, where national gold OA publications represent 63% of the total of gold OA publications.

Discussion

The purpose of this study is to present a global view of the state of Open Access penetration at the institutional level. For this, we have included all universities appearing in the 2019th edition of the Leiden Ranking and retrieved all their publications from Web of Science. These have been crossed with Unpaywall, a database which identifies evidences of OA for publications under the requirement that they have a DOI assigned to them. An important limitation of this tool is that it is dependent on DOIs, which means that we underestimate OA penetration overall, and especially in the Arts and Humanities fields (Gorraiz et al., 2016). Based on evidences of OA presence, we classified OA publications into four types: gold, green, hybrid and bronze. Overall, we find that around 41% of all publications contained in our data set are openly accessible. Green OA is the most common type of OA (77%), followed by Gold OA (33%).

Still, we find great differences between countries. For instance, Brazilian universities show a higher median share of Gold OA than Green OA, being the only case where this happen. Probably, the strong commitment with national OA publishing via the SciELO programme is behind such trend (Meneghini, Mugnaini & Packer, 2006). United Kingdom, Netherlands, Austria and Sweden show similar levels of gold and hybrid OA, a surprising pattern as the levels of OA awareness and the types of mandates implemented in these countries is quite different (Schmidt & Kuchma, 2012). These differences between countries are observed also at the continental level (Figure 6) with Europe leading on OA penetration, followed by North America, and Asia and Africa lagging behind. However, it also yields many differences between universities from the same region, with only universities from Oceania and South America showing similar ratios of OA presence.

A closer look into green OA reveals some counterintuitive findings. First, the presence of repositories such as PubMed Central (PMC) which, although laudable, distort to some extent our perception of what is green OA and what it is not, particularly at the institutional level. This repository (and there might be others), indexes automatically OA literature, meaning that it includes self-archived publications as well as those from OA journals and OA publications from toll journals (Hybrid OA). Depending on how restrictive we are on our definition of green OA (i.e., self-archived by the author), we might disregard this source and hence reduce the overall presence of this type of OA. This along with the inclusion of bronze OA, evidence some

discrepancies between the conceptual definition of OA and how it is operationalized in practice, leading the way to alternative conceptual framings of OA which might be closer to actual evidence of OA (e.g., Martín-Martín et al., 2018a). Here, we propose looking into the share of publications stored in universities' own repository and highlight some cases of good practices such as Bilkent University or City University London (Table 1).

In the case of gold OA, where the definition is much clearer, the intrusion of an author pays model (or APC model), along with the emergence of predatory journals (Grudniewicz et al., 2019), has led the way to much criticism as to the quality of OA journals (Bohannon, 2013). While it is out of the scope of this study to analyse or compare the quality of OA journals, we do attempt to characterize such journals. For this, we expand on the modelling proposed by Torres-Salinas, Robinson-Garcia & Moed (2019), and use three variables to characterize countries' gold OA publishing: language of publication, journals' editing country and the inclusion of an APC model (Figure 9). This way we can identify outliers following alternative models of publishing (such as the aforementioned case of Brazil), evidencing that in some cases, publishing in OA journals is more related with other factors, such as publishing in national journals or non-English language rather than with the fact that the journal is offered in Open Access.

All in all, this paper presents a first attempt at analysing OA presence at the institutional level. While the study is descriptive in nature, it opens the opportunity for institutions, funding agencies and national science policy officers to better understand the expansion of OA in their country and better design and model effectives mandates of OA. Furthermore, new indicators can be designed which may fit into indicator frameworks of Open Science (Schomberg et al., 2019), moving away from metrics of excellence to metrics of openness and transparency.

Acknowledgements

The authors would like to thank Henri de Winter for technical support and Jason Priem and Heather Piwowar, developers of Unpaywall, for fruitful discussions on the identification of OA types.

References

Aguillo IF, Ortega JL, Fernández M, Utrilla AM. 2010. Indicators for a webometric ranking of open access repositories. *Scientometrics* 82:477–486. DOI: 10.1007/s11192-010-0183-y.

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

Arlitsch K, Grant C. 2018. Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape. *Journal of Library Administration* 58:264–281. DOI: 10/gf4w3g.

Bohannon J. 2013. Who's Afraid of Peer Review? *Science* 342:60–65. DOI: 10.1126/science.342.6154.60.
David PA. 2008. The Historical Origins of "Open Science": An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. *Capitalism and Society* 3. DOI: 10/fnrmrn.
Else H. 2018a. Radical open-access plan could spell end to journal subscriptions. *Nature* 561:17. DOI: 10.1038/d41586-018-06178-7.
Else H. 2018b. Funders flesh out details of Europe's bold open-access plan. *Nature*. DOI: 10.1038/d41586-018-07557-w.
Gorraiz J, Melero-Fuentes D, Gumpenberger C, Valderrama-Zurián J-C. 2016. Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics* 10:98–109. DOI: 10/f8d2bp.
Gorraiz J, Wieland M. 2009. Multi-authored publications: their influence in the distribution of the financing costs in world licenses. *Research Evaluation* 18:215–220. DOI: 10.3152/095820209X466919.
Gorraiz J, Wieland M, Gumpenberger C. 2012. Bibliometric practices and activities at the University of Vienna. *Library Management* 33:174–183. DOI: 10.1108/01435121211217199.
Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, Ardern C, Balcom L, Barros T, Berger M, Ciro JB, Cugusi L, Donaldson MR, Egger M, Graham ID, Hodgkinson M, Khan KM, Mabizela M, Manca A, Milzow K, Mouton J, Muchenje M, Olijhoek T, Ommaya A, Patwardhan B, Poff D, Proulx L, Rodger M, Severin A, Strinzel M, Sylos-Labini M, Tamblyn R, Niekerk M van, Wicherts JM, Lalu MM. 2019. Predatory journals: no definition, no defence. *Nature* 576:210–212. DOI: 10/dgtx.
Harnad S. 2007. Mandates and metrics: How open repositories enable universities to manage, measure and maximise their research assets.
Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Oppenheim C, Hajjem C, Hilf ER. 2008. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review* 34:36–40. DOI: 10.1080/00987913.2008.10765150.
Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Oppenheim C, Stamerjohanns H, Hilf ER. 2004. The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review* 30:310–314. DOI: 10/gdkw3j.
van Leeuwen TN, Tatum C, Wouters PF. 2018. Exploring possibilities to use bibliometric data to monitor gold open access publishing at the national level. *Journal of the Association for Information Science and Technology* 69:1161–1173. DOI: 10.1002/asi.24029.
van Leeuwen TN, Costas R, Robinson-Garcia N. 2019.Indicators of open access publishing in the CWTS Leiden Ranking 2019. *Available at https://www.cwts.nl/blog?article=n-r2w2a4&title=indicators-of-open-access-publishing-in-the-cwts-leiden-ranking-2019*
Martín-Martín A, Costas R, Leeuwen TN van, López-Cózar ED. 2018a. Unbundling Open Access dimensions: a conceptual discussion to reduce terminology inconsistencies. DOI: 10.17605/osf.io/7b4aj.

476 Martín-Martín A, Costas R, van Leeuwen T, Delgado López-Cózar E. 2018b. Evidence of open
477 access of scientific publications in Google Scholar: A large-scale analysis. *Journal of*
478 *Informetrics* 12:819–841. DOI: 10.1016/j.joi.2018.06.012.
479 Meneghini R, Mugnaini R, Packer AL. 2006. International versus national oriented Brazilian
480 scientific journals. A scientometric analysis based on SciELO and JCR-ISI databases.
481 *Scientometrics* 69:529–538. DOI: 10.1007/s11192-006-0168-z.
482 Moedas C. 2015. *Open Innovation, Open Science, Open to the World*. European Commission's
483 Directorate-General for Research & Innovation (RTD).
484 Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J,
485 Haustein S. 2018. The state of OA: a large-scale analysis of the prevalence and impact of Open
486 Access articles. *PeerJ* 6:e4375. DOI: 10.7717/peerj.4375.
487 Robinson-Garcia N, Costas R, van Leeuwen TN. 2019. Indicators of Open Access for
488 universities. *arXiv:1906.03840 [cs]*.
489 Schmidt B, Kuchma I. 2012. *Implementing Open Access Mandates in Europe: OpenAIRE Study*
490 *on the Development of Open Access Repository Communities in Europe*. Universitätsverlag
491 Göttingen.
492 Schomberg R von, Holbrook BJ, Oancea A, Kamerlin SCL, Ràfols I, Jacob M, Wouters P. 2019.
493 *Indicator frameworks for fostering open knowledge practices in science and scholarship.*
494 Brussels: European Commission's Directorate-General for Research & Innovation (RTD).
495 Torres-Salinas D, Robinson-Garcia N, Moed HF. 2019. Disentangling Gold Open Access. In:
496 Glänzel W, Moed HF, Schmoch U, Thelwall M eds. *Handbook of Science and Technology*
497 *Indicators*. Cham: Springer, 129–144.
498 Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, van Eck NJ, van Leeuwen
499 TN, van Raan AFJ, Visser MS, Wouters P. 2012. The Leiden ranking 2011/2012: Data
500 collection, indicators, and interpretation. *Journal of the American Society for Information*
501 *Science and Technology* 63:2419–2432. DOI: 10.1002/asi.22708.
502