Open Access uptake by universities worldwide

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

1 Delft Institute of Applied Mathematics, Delft University of Technology, Delft, Netherlands

2 Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands

3 Centre for Research on Evaluation, Science and Technology, University of Stellenbosch, Stellenbosch, South Africa

Corresponding Author: Nicolas Robinson-Garcia
Email address: n.robinsongarcia@tudelft.nl

The implementation of policies promoting the adoption of an open science (OS) culture must be accompanied by indicators that allow monitoring the uptake of such policies and their potential effects on research publishing and sharing practices. This study presents indicators of Open access (OA) 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 OA and OS practices and policies.

1 Open Access uptake by universities worldwide

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

5 [1] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, Netherlands

6 [2] Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands

7 [3] Centre for Research on Evaluation, Science and Technology, University of Stellenbosch,
8 Stellenbosch, South Africa

10 Corresponding Author:

11 Nicolas Robinson-Garcia[1]

12 Building 28, Mourik Broekmanweg 6, 2628 XE Delft, Netherlands

13 Email address: elrobinster@gmail.com

15 Abstract

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

25 Introduction

26 The implementation of policies promoting the adoption of an open science (OS) culture must be
27 accompanied by indicators that allow monitoring the penetration of such policies and their
28 potential effects on research publishing and sharing practices. In this paper we present Open
29 access (OA) indicators for universities worldwide. We analyse the presence of OA by type of
30 access, field differences and comparisons with scientific impact and international collaboration.

31   We explore discrepancies between the operationalization of OA indicators and the different
32   definitions of OA provided in the literature.
33
34   The notion of OS goes back to the sixteenth Century (David, 2008), but it has recently gained
35   relevance as the European Union (EU) introduced it as a pivotal stone in their research
36   programmes (Moedas, 2015). Within the different directives set up to achieve it, OA has become
37   one of the first milestones. Initiatives such Plan S (Else, 2018a,b) or the European Commission's
38   Open Science Monitor[1] exemplify such efforts and the prioritization of OA for these agencies.
39   The latter being the tool the European Commission is using to monitor its uptake. However,
40   more granular levels of analysis are needed to better understand how OA is expanding, which
41   OA models are being implemented and what the potential side-effects of such models are.
42   Universities have been supporting OA for many years now. The most common strategy has been
43   to build and maintain institutional repositories, and introduce mandates that oblige their
44   researchers to deposit pre or postprints of their publications (Harnad, 2007; Harnad et al., 2008).
45   There is also evidence of institutions promoting OA publications by sponsoring the costs derived
46   from the article processing charges (APC) of open journals (Gorraiz & Wieland, 2009; Gorraiz,
47   Wieland & Gumpenberger, 2012). In most cases, institutions are faced with the challenge of
48   determining the success of such initiatives and monitoring the compliance of their researchers
49   with international and national OA mandates. Initiatives such as the ranking of OA repositories
50   (Aguillo et al., 2010) offer partial information on the share of OA availability at the institutional
51   level, as they only provide details on the among of documents stored in institutional repositories,
52   irrespective of the overall research output. Although valuable, it is still insufficient, as
53   institutional repositories may not be the main vehicle used by researchers to make their outputs
54   openly accessible (Arlitsch & Grant, 2018), and not all researchers, even at the same university,
55   comply with their institutional mandates in the same manner.
56
57   Until recently, there were no more than estimates as to the amount of OA publications. However,
58   the development of platforms like CrossRef, DOAJ or even Google Scholar, along with
59   computational advancements on web scrapping, have led to a plethora of large-scale analyses to
60   empirically identify OA literature (Archambault et al., 2014; van Leeuwen, Tatum & Wouters,
61   2018; Piwowar et al., 2018; Martín-Martín et al., 2018b). Overall, these studies report that
62   around half of the scientific literature is freely available but point towards the increasing
63   availability of publications which do not adhere strictly to what is considered OA. The game
64   changer in this respect, has been Unpaywall (Piwowar et al., 2018), a product developed by the
65   non-profit Our Research[2], which tracks OA versions of published research with a document

---

[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/

66 identifier (e.g., DOI), including e-prints self-archived in repositories, recently becoming the most
67 standard mechanism to identify OA.
68
69 In this paper we present a first attempt at analyzing OA at the institutional level. The goal main
70 of the study is to provide methodological insights that can ease the analytical use and
71 interpretation of OA indicators while providing a general overview of institutional OA uptake.
72 Hence, the paper is in nature descriptive, not aiming at responding specific research questions on
73 OA uptake, but to set the basis so that more specific research questions can be developed in the
74 future in an informed way. In order to achieve the proposed goal, we structure our findings in the
75 following way. First, we inform on how OA is being achieved in different institutions and
76 countries, describe national trends, and pathways by which OA is being expanded. Second, we
77 deepen the analysis into green and gold OA types to understand potential discrepancies between
78 the common understanding of these two types OA and how it is actually operationalized in
79 Unpaywall. The results of this study have been recently incorporated to the 2019 edition of the
80 Leiden Ranking released in May 2019 (van Leeuwen, Costas & Robinson-Garcia, 2019) and a
81 first version was presented at the ISSI 2019 Conference (Robinson-Garcia, Costas & van
82 Leeuwen, 2019).
83

84 ## Materials & Methods

85 In this paper we use different sets of data sources and combine different methods to determine
86 OA. Publication data is retrieved from the CWTS in-house version of the Web of Science.
87 Unlike in the Leiden Ranking (restricted to article and reviews), here we report indicators for
88 letters, articles and reviews (that is documents considered *citable*) indexed in the Science
89 Citation Index Expanded, Social Sciences Citation Index and Arts & Humanities Citation Index
90 for the 2014-2017 period. We link publications to the 963 universities identified in the Leiden
91 Ranking database via their disambiguated list of institutional names, also hosted at CWTS
92 (Waltman et al., 2012). Publications are assigned to five fields of science, following the
93 methodology employed in the Leiden Ranking[3]. These fields are: Biomedical and Health
94 Sciences, Life and Earth Sciences, Mathematics and Computer Science, Physical Sciences &
95 Engineering, and Social Sciences and Humanities. The supplementary data offers indicators
96 aggregated at the institutional level as detailed record metadata is subject to proprietary rights
97 and cannot be openly shared.
98 For each publication, we identify if they are openly accessible and the type of OA by querying
99 the Unpaywall information. A dump version of the Unpaywall database retrieved in April, 2019.
100 Unpaywall relies on Digital Object Identifiers (DOI), which means that we will only include
101 records which have a DOI assigned to them. Furthermore, the Unpaywall API does not label

---

[3] A detailed description the assignment of publications to fields is provided here
https://www.leidenranking.com/information/fields. It is important to note here that humanities journals
publications are taken out of the Leiden Ranking publication set.

102 types of OA but records different pieces of evidence of OA availability of each publication.
103 More information on the Unpaywall approach to OA is available at their User Guide offered for
104 researchers (http://unpaywall.org/data-format).
105

106   [Figure 1 here]

107

108 Four types of OA are considered. These four types of OA are defined as follows:
109

110   • **Green OA.** Self-archived versions of a manuscript. Here the responsibility lies on the
111   authors of the publication, or institutional colleagues such as central library staff
112   members, who oversee depositing the document in a repository. This version of the
113   document may not correspond with the final version of the publisher.
114   • **Gold OA.** This refers to journals which publish all their manuscripts in OA regardless of
115   the business model they follow (e.g., publicly sponsored, author pays).
116   • **Hybrid OA**. Toll access (non-OA) journals make specific publications openly accessible
117   usually after the author pays a fee, claiming an alleged need to account for potential
118   losses derived from subscription fees.
119   • **Bronze OA.** This OA type was first suggested by Piwowar et al. (2018) and refers to
120   free-to-read articles made available by publishers, without an explicit mention to any OA
121   license. This OA is not subjected to copyright conditions set to be defined as OA (i.e.,
122   they do not ensure perpetual free access).
123

124 The labelling of OA types is described in Figure 1 and already highlights some of the difficulties
125 raised when trying to define what is actually OA (Torres-Salinas, Robinson-Garcia & Moed,
126 2019). The Unpaywall database provides a set of different pieces of OA evidence for each DOI.
127 For each piece of evidence, we study all the metadata labels referring to the OA status of the
128 publication. Thus, when one piece of evidence suggests that a paper belongs to an OA journal
129 (gold OA), this automatically overrides bronze or hybrid OA. As observed in Figure 1, gold,
130 bronze and hybrid are conceptually incompatible types of OA, as these documents are all
131 provided by the publishers, distinguishing themselves by the type of journal (OA or toll) or the
132 presence of an OA license (hybrid or bronze). The only exception made with green OA, which
133 could overlap with any of the other three types.
134

135 Overall, a total of 4,621,721 distinct publications records are examined. 4,620,666 include a
136 DOI, out of which 1,881,193 records were identified as OA. Figure 2 shows how these OA
137 publications are distributed by type. 77% of all OA publications were green OA, followed by
138 gold OA (33%), bronze OA (20%) and hybrid OA (16%). However, there is a substantial overlap
139 between each of these latter OA types and green OA. 81% of all gold OA publications are also in
140 green OA, for hybrid the share which is also green is 63%, and of hybrid are 45% for bronze OA.

141
142   [Figure 2 here]
143

144 The results are reported at different levels. First analyses investigate the share of OA on the
145 overall output of each university, differences by country, continent and field. We then look
146 specifically into the two main OA types: green and gold. We focus specifically on these two
147 types as they represent the largest groups of OA publications. In the case of green OA
148 publications, we focus on two aspects. First, the relation between green OA documents produced
149 by an institution and share of which are stored in their institutional repository. Second, potential
150 distortions on how OA types are operationalized and how they are commonly defined,
151 specifically looking into the role played by PubMed Central (PMC). In the case of gold OA, we
152 introduce different national models of gold OA publishing. We characterize gold OA publishing
153 based on three variables: share of papers published in national journals, share of papers
154 published in English language and share of papers published in journals following including
155 APCs. Language of documents and journal's country are identified using data from Web of
156 Science. In the case of the latter, we identify the country of the journal by querying the field
157 Publisher Address (PA). This approach is not exempt of limitations, as some publishers have an
158 international outlook, while others may shape their geographic focus based on the location of
159 their editors. Hence, the results of this analysis should be interpreted with caution and as a first
160 attempt for developing taxonomies of gold OA publishing.
161
162 In the case of APCs, we queried the Directory of Open Access Journals (DOAJ). Here we must
163 note that this is not a comprehensive list of OA journals. Unpaywall identifies a larger number of
164 gold OA journals (n= 11,601) than DOAJ (n= 11,365), and for which we have no information on
165 APCs. Therefore, the numbers on gold OA journals with/out APCs provided represent a lower
166 bound of all the gold OA journals for which APC information is available via DOAJ. A total of
167 768 APC journals were identified. After some inspection, we found some inconsistencies in the
168 way APC is defined according to DOAJ. That is, not in all cases, APC refers to an author pays
169 model, but in some cases, journals offer an optional subscription fee for those interested on
170 accessing to printed versions of the journal. This is the case for many journals stored in the
171 SciELO platform which are free of costs for both readers and authors, but which offers the
172 option to pay a subscription fee for printed versions of the journal.
173

Results

General overview

176 In Figure 3 we consider the proportion of OA publications by countries. Only countries with at
177 least 10 universities listed in the Leiden Ranking are shown. The median share of publications
178 openly available of universities worldwide is 43%. British universities have by far the largest
179 share of OA publications (median=74%), followed by Sweden (median=56%) and Austria

180   (median=54%). Except for the United States (median=51%) and Brazil (median=47%), all
181   countries above world median are European. Asian countries, as well as Canada and Australia
182   show OA shares below the world median.
183

184   [Figure 3 here]

185

186   We disaggregate by type of OA in Figure 4. Most OA publications are openly accessible via the
187   green route, and hence the similarity between Figure 3 and Figure 4A. In the case of gold OA
188   (Figure 4B) a very different image is seen. Brazilian universities stand out with a median of 30%
189   publications in Gold OA. Sweden is placed in second, along with Taiwan (median=18% for both
190   countries). Universities from United Kingdom (median=17%), Austria (median=15%) and the
191   Netherlands (median=13%) correspondingly, show the highest share of hybrid OA publications.
192   While for bronze OA, it is universities from Japan (median=15%), Turkey (median=13%) and
193   the Netherlands (median=12%) that stand out.
194

195   [Figure 4 here]

196

197   Figure 5 shows the predominance of each OA type by field and at the university level (each point
198   represents the share of a university in each field and type of OA grouping). The average share of
199   OA publications is 42.8%. The highest median is found in the Biomedical and Health Sciences
200   (49.1%), while Social sciences and Humanities exhibit the lowest shares of OA (36.5%). Green
201   OA is the most predominant form of OA regardless of the field (median of 33.2% in the 'All
202   sciences' group). Again, the largest average is found in Biomedical and Health Sciences (39.0%)
203   and the lowest in Social sciences and Humanities (28.0%). Overall, universities publish on
204   average 14.7% of their publications in OA journals. For Biomedical & Health Sciences the
205   average increases up to 19.3%, while in Mathematics & Computer Science it drops to 9.0%. In
206   the case of Hybrid OA, an average of 7.1% of papers in universities are published under this
207   modality. This figure increases in the case of Physical Sciences and Engineering to 7.9%, while
208   in Social sciences & Humanities it represents an average of 4.6% of the output. Bronze OA,
209   although it not strictly OA as it does not ensure sustainable access, is more common on average
210   than Hybrid OA, with an overall share of 8.5% which goes up to 11.1% for Biomedical & Health
211   Sciences, but with a presence on average of 3.7% in Mathematics & Computer Science.

212

213   [Figure 5 here]

214

215   We also note large differences by geographical region (Figure 6). Europe (50.1%) and North
216   America (49.1%) are the continents with the universities sharing the largest proportions of their
217   output in OA. In the other extreme we find Asia (32.5%) and Africa (39.1%). In the former two
218   continents, green OA is by large the most common OA type (41.1% in Europe and 40.6% in
219   North America) with gold OA lagging behind by far as the second option (15.4% and 12.0%
220   respectively). In South America, median shares of green (29.2%) and gold OA (27.0%) by
221   university are practically identical. Shares of hybrid and bronze OA are on median below 10%
222   for all continents except for bronze OA in North America (11.2%).
223

224   [Figure 6 here]

225

226   **University profiling**

227   It is remarkable that differences between and within universities can be quite significant. In
228   Figure 7 we take a closer look into the disciplinary profile of a set of universities based on the
229   type and proportion of OA output by field. To illustrate the OA institutional profiling of
230   universities, we use radar charts and select in each row the three universities with the largest
231   output (considering their full counting) in North America, Europe, Africa, South America and
232   Asia, respectively. In the first row, we observe the three largest universities in North America,
233   two from the United States and one from Canada. The two US universities have above half of
234   their output in green OA, with Social sciences and humanities, just below the 50% threshold. In
235   the case of the University of Toronto, the shares are much lower, ranging between 39% green
236   OA in Biomedical and Health Sciences and 3% bronze OA in Mathematics and Computer
237   Science. The three largest universities in Europe are all from the United Kingdom. Again, green
238   OA is clearly the most common OA option in all fields within these universities, showing more
239   homogeneity across the three institutional profiles. However, Social Sciences and Humanities
240   tend to have lower shares for the universities of Cambridge and Oxford than for University
241   College London.
242

243   [Figure 7 here]

244

245   Regarding Africa (third row), two of the three universities showcased are South African, while
246   the third one is Egyptian. In the case of Cairo University, no OA type in any field reaches a
247   quarter of the total output of the university. In the other two cases, the profiles are quite similar,
248   with the University of Cape Town exhibiting higher shares of green OA than the University of
249   Witwatersrand. For South America, three Brazilian universities outstand as the largest ones;
250   Universidade de São Paulo, Universidade Estadual Paulista and Universidade Estadual de
251   Campinas. The gold OA preponderance previously observed at an aggregate level both for the

252 continent and Brazil, is also noted at the institutional level in all three universities. However, we
253 do observe that such preponderance is coming mainly from the Life and Earth Sciences and the
254 Social Sciences and Humanities. Finally, for Asia (last row), we profile three Chinese
255 universities for which green and gold OA shares go hand in hand in all three cases, with the
256 exception of the field of Biomedical and Health Sciences, where green OA reaches higher shares
257 of the total output.
258

259 **Green Open access and self-archiving**

260 We will now delve into green OA, to better comprehend the indicators shown displayed on this
261 typology. Green OA was originally defined as self-archiving of preprint or post print versions of
262 published manuscripts. That means that green OA is achieved as the result of a proactive attitude
263 of the authors or an institutional colleague, like librarians, towards OA. In their seminal paper,
264 Harnad et al. (2004) go beyond such definition, and indicate that "the self-archiving method with
265 the greatest potential to provide OA is self-archiving in the author's own university's OAI-
266 compliant Eprint Archives" (p. 312). Hence, one could expect to see in the green OA indicators,
267 shares of institutional self-archiving of a university's output. However, a closer look into what is
268 considered as green following the identification procedure used based on Unpaywall data, shows
269 that this is not the case for two reasons.
270

271 First, the assignment of OA output to each university is given based the affiliation of authors and
272 not the contents of institutional repositories. This means that universities with large proportions
273 of their output in green OA may not be succeeding on storing their output in their institutional
274 repositories themselves. Table 1 shows the top 20 universities with the largest shares of their
275 output in green OA. Along with the total number of publications and green OA publications, we
276 provide a threshold of the share of publications which are stored in their own institutional
277 repository. We identify the lower band of the threshold by individually querying the URL string
278 of each university's repository. The upper band results from also including URL string
279 containing hdl.handle.net, which is the URL used when linking through the HANDLE identifier,
280 a similar identifier to DOIs but assigned by repositories. Some universities do have most of their
281 output accessible thanks to their own institutional repositories. For instance, 98% of green OA
282 publications from Bilkent University are stored in their own repository.
283

284 [Table 1 here]
285

286 Second, low coverages of green OA output in institutional repositories can be due to inter-
287 institutional collaboration (i.e. collaboration with other institutional partners that apply more
288 systematic archiving policies) or self-archiving in thematic (e.g., ArXiv) or supranational
289 repositories (e.g., Zenodo). However, there is a second phenomenon which drifts further away
290 the original definition of green OA from the actual numbers that are reported based on the

291 general labelling obtained via Unpaywall. That is, the effect of repositories which store OA
292 documents without authors' intervention. Previously, we referred to this as different perspectives
293 of green OA based on "the degree of engagement" of the authors (van Leeuwen, Costas &
294 Robinson-Garcia, 2019). We distinguish between two perspectives: 1) self-archiving, defined as
295 the deliberate action of an author or librarian to archive publications in a repository, and 2)
296 general archiving, where the archival function is still taking place, but without the explicit
297 intervention of the author or librarian. So far, we have identified one macro repository following
298 this general archiving perspective; PubMed Central (https://www.ncbi.nlm.nih.gov/pmc/). This
299 source alone represents 60.8% of the green OA literature identified. However, some of its
300 contents are retrieved from elsewhere, including OA journals such as Plos ONE. 86.5% of the
301 881,834 documents in PMC are simultaneously also gold, bronze or hybrid OA. The remaining
302 13.5% is accessible via another repository as well as PMC. As it is indeed a repository, in this
303 study it is considered as a green OA source, but the effect of such decision in OA shares at the
304 institutional level is highly significant. Figure 8 shows the effect of PMC on the shares of green
305 OA. 49 universities are shown, these are those for which green OA deposited in PMC represents
306 95% or more of their total number of green OA publications. While self-archived and PMC
307 publications can overlap (as more than one instance of OA evidence can be found per
308 publication), in some cases the difference between defining PMC publications as green or not
309 can derive on up to more than 10,000 publications, as in the case of University of Texas,
310 Houston.
311
312   [Figure 8 here]
313   [Table 2 here]
314
315 In Table 2 we aggregate the set of universities at the country level, to identify in which countries
316 the inclusion of PMC as green OA affects the most their figures. The greatest effect is observed
317 in Taiwan (65.5% of their total green OA), South Korea (62.2%) and China (60.1%).
318 Furthermore, we observe that most of the documents coming from PMC are provided through
319 another OA route, mostly gold, but also hybrid and bronze. This shows again the introduction of
320 some degree of duplication of other OA types into green when including PMC and how the way
321 we define and operationalize each of the OA types can affect the final numbers provided.
322

Gold Open access models

324 As previously observed, Gold OA is the second largest type of OA of the four analysed here
325 (Figure 2), but with some notable exceptions like the case of Brazil (Figure 4B). Torres-Salinas
326 et al. (2019) highlight three models to characterize gold OA publishing from their analysis on
327 Gold OA. The first one represents countries which publish in OA journals from big publishing
328 firms and with a high Journal Impact Factor. Countries like United Kingdom, Germany or the
329 Nordic countries fit into this model. A second model showcases countries publishing in national
330 low Impact Factor OA journals, such as Brazil or India. The third model is a combination of the

331   previous two, where they point out at countries like Poland or Spain. In Figure 9 we take a
332   similar approach looking at three variables for Gold OA publishing: share of gold OA
333   publications in APC journals, share of gold OA publications in English language and share of
334   gold OA publications from national journals. We observe that patterns are quite stable for the
335   three variables. Most countries publish up to 25% of their output in national OA journals. APCs
336   are paid for a range between 50% and 75% of their gold OA publications, and almost all of it is
337   published in English language.
338
339   [Figure 9 here]
340
341   This pattern is followed by most countries, but some differences can be observed. For instance,
342   United States and United Kingdom represent countries with high level of APC publishing, high
343   shares in national language and almost exclusively in English language. Switzerland also fits into
344   this pattern despite being a non-English speaking country. Another differing pattern is observed
345   for countries like Spain or Portugal, where the share of English language publications is much
346   lower although the share of national publications is still below 25% (23% and 21% respectively).
347   In the case of Poland, although 98% of Gold OA publications are in English language, 45%
348   come from national journals, with APC publications in the lower bound of the 50%-75% interval
349   (52%). A similar pattern is followed by South Africa. Finally, we highlight the case of Brazil,
350   where national gold OA publications represent 63% of the total of gold OA publications.
351

Discussion

353   The purpose of this study is to present a global view of the state of OA uptake at the institutional
354   level. For this, we have included all universities appearing in the 2019th edition of the Leiden
355   Ranking and retrieved all their publications from Web of Science. These have been crossed with
356   Unpaywall, a database which identifies evidences of OA for publications under the requirement
357   that they have a DOI assigned to them. An important limitation of this tool is that it is dependent
358   on DOIs, which means that we underestimate OA penetration overall, and especially in the Arts
359   and Humanities fields (Gorraiz et al., 2016). Based on evidences of OA presence, we classified
360   OA publications into four types: gold, green, hybrid and bronze. Overall, we find that around
361   41% of all publications contained in our data set are openly accessible. Green OA is the most
362   common type of OA (77%), followed by Gold OA (33%).
363
364   Still, we find great differences between countries. For instance, Brazilian universities show a
365   higher median share of Gold OA than Green OA, being the only case where this happen. This is
366   a paradigmatic case, arguably the result of a long-standing OA policy commitment promoting
367   national OA journals via the SciELO programme (Meneghini, Mugnaini & Packer, 2006).
368   Furthermore, it goes beyond Brazilian universities and includes other South American countries
369   such as Colombia or Chile (Packer, 2009; Minniti, Santoro & Belli, 2018), not necessarily fully
370   covered in this study, due to the restrictions on the set of institutions included (only those present

371 in the Leiden Ranking). United Kingdom, Netherlands, Austria and Sweden show similar levels
372 of gold and hybrid OA, a surprising pattern as the levels of OA awareness and the types of
373 mandates implemented in these countries is quite different (Schmidt & Kuchma, 2012). United
374 Kingdom exhibit a strong OA uptake as a result of the implementation of policies which
375 reinforce especially, the green route (Chan, 2019). In this sense, the UK's Research Evaluation
376 Framework plays an important role on promoting OA, as any publication submitted for
377 evaluation are required be openly accessible (Hatzipanagos & Gregson, 2014). These differences
378 between countries are observed also at the continental level (Figure 6) with Europe leading on
379 OA penetration, followed by North America, and Asia and Africa lagging behind. However, it
380 also yields many differences between universities from the same region, with only universities
381 from Oceania and South America showing similar ratios of OA presence.
382
383 A closer look into green OA reveals some counterintuitive findings. First, the presence of
384 repositories such as PubMed Central (PMC) which, although laudable, distort to some extent our
385 perception of what is green OA and what it is not, particularly at the institutional level. This
386 repository (and there might be others), indexes automatically OA literature, meaning that it
387 includes self-archived publications as well as those from OA journals and OA publications from
388 toll journals (Hybrid OA). Depending on how restrictive we are on our definition of green OA
389 (i.e., self-archived by the author), we might disregard this source and hence reduce the overall
390 presence of this type of OA. This along with the inclusion of bronze OA, evidence some
391 discrepancies between the conceptual definition of OA and how it is operationalized in practice,
392 leading the way to alternative conceptual framings of OA which might be closer to actual
393 evidence of OA (e.g., Martín-Martín et al., 2018a). Here, we propose looking into the share of
394 publications stored in universities' own repository and highlight some cases of good practices
395 such as Bilkent University or City University London (Table 1).
396
397 In the case of gold OA, where the definition is much clearer, the intrusion of an author pays
398 model (or APC model), along with the emergence of predatory journals (Grudniewicz et al.,
399 2019), has led the way to much criticism as to the quality of OA journals (Bohannon, 2013). In
400 the present study, the study of the presence of predatory journals in the counts of OA
401 publications at the institutional-level is not directly approached, and we estimate that the study of
402 this aspect may be restricted by the fact that we are only considering publications covered in the
403 Web of Science database (from which we normalize institutional names and extract scientific
404 fields), which is a database with a more restrictive coverage of scientific publications. But a
405 similar approach to the one presented here could be performed to larger databases (such as
406 Dimensions or Microsoft Academic Graph) in which predatory journals could have a stronger
407 presence (an aspect which remains to a large extent unclear, and requires additional research).
408
409 While it is out of the scope of this study to analyse or compare the quality of OA journals, we do
410 attempt to characterize such journals. For this, we expand on the modelling proposed by Torres-

411 Salinas, Robinson-Garcia & Moed (2019), and use three variables to characterize countries' gold
412 OA publishing: language of publication, journals' editing country and the inclusion of an APC
413 model (Figure 9). This way we can identify outliers following alternative models of publishing
414 (such as the aforementioned case of Brazil), evidencing that in some cases, publishing in OA
415 journals is more related with other factors, such as publishing in national journals or non-English
416 language rather than with the fact that the journal is offered in OA.
417

Conclusion

419 While the study is descriptive in nature, it opens the opportunity for institutions, funding
420 agencies and national science policy officers to better understand the expansion of OA in their
421 country and better design and model effectives mandates of OA. Furthermore, new indicators
422 can be designed which may fit into indicator frameworks of OS (Schomberg et al., 2019),
423 moving away from metrics of excellence to metrics of openness and transparency.
424

Acknowledgements

426 The authors would like to thank Henri de Winter for technical support, Jason Priem and Heather
427 Piwowar, developers of Unpaywall, for fruitful discussions on the identification of OA types,
428 and David Moher and Stefanie Haustein who reviewed this manuscript and improved it with
429 their suggestions and comments. The opinions expressed in this study reflect only the author's
430 view. The European Commission is not responsible for any use that may be made of the
431 information it contains
432

References

434 Aguillo IF, Ortega JL, Fernández M, Utrilla AM. 2010. Indicators for a webometric ranking of open
435   access repositories. *Scientometrics* 82:477–486. DOI: 10.1007/s11192-010-0183-y.
436 Archambault É, Amyot D, Deschamps P, Nicol A, Provencher F, Rebout L, Roberge G. 2014. Proportion
437   of open access papers published in peer-reviewed journals at the European and world levels—1996–
438   2013. https://digitalcommons.unl.edu/scholcom/8/
439 Arlitsch K, Grant C. 2018. Why So Many Repositories? Examining the Limitations and Possibilities of
440   the Institutional Repositories Landscape. *Journal of Library Administration* 58:264–281. DOI:
441   10.1080/01930826.2018.1436778.
442 Bohannon J. 2013. Who's Afraid of Peer Review? *Science* 342:60–65. DOI:
443   10.1126/science.342.6154.60.
444 Chan TT. 2019. *Open research policies in the United Kingdom - Open Science Monitor Case Study*.
445   Brussels: Directorate-General for Research and Innovation (European Commission).
446 David PA. 2008. The Historical Origins of "Open Science": An Essay on Patronage, Reputation and
447   Common Agency Contracting in the Scientific Revolution. *Capitalism and Society* 3. DOI: 10/fnrmrn.
448 Else H. 2018a. Radical open-access plan could spell end to journal subscriptions. *Nature* 561:17. DOI:
449   10.1038/d41586-018-06178-7.

450 Else H. 2018b. Funders flesh out details of Europe's bold open-access plan. *Nature*. DOI:
451   10.1038/d41586-018-07557-w.
452 Gorraiz J, Melero-Fuentes D, Gumpenberger C, Valderrama-Zurián J-C. 2016. Availability of digital
453   object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics* 10:98–109. DOI:
454   10/f8d2bp.
455 Gorraiz J, Wieland M. 2009. Multi-authored publications: their influence in the distribution of the
456   financing costs in world licenses. *Research Evaluation* 18:215–220. DOI: 10.3152/095820209X466919.
457 Gorraiz J, Wieland M, Gumpenberger C. 2012. Bibliometric practices and activities at the University of
458   Vienna. *Library Management* 33:174–183. DOI: 10.1108/01435121211217199.
459 Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, Ardern C, Balcom L, Barros T,
460   Berger M, Ciro JB, Cugusi L, Donaldson MR, Egger M, Graham ID, Hodgkinson M, Khan KM,
461   Mabizela M, Manca A, Milzow K, Mouton J, Muchenje M, Olijhoek T, Ommaya A, Patwardhan B,
462   Poff D, Proulx L, Rodger M, Severin A, Strinzel M, Sylos-Labini M, Tamblyn R, Niekerk M van,
463   Wicherts JM, Lalu MM. 2019. Predatory journals: no definition, no defence. *Nature* 576:210–212. DOI:
464   10/dgtx.
465 Harnad S. 2007. Mandates and metrics: How open repositories enable universities to manage, measure
466   and maximise their research assets.
467 Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Oppenheim C, Hajjem C, Hilf ER.
468   2008. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials
469   Review* 34:36–40. DOI: 10.1080/00987913.2008.10765150.
470 Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Oppenheim C, Stamerjohanns H, Hilf
471   ER. 2004. The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review*
472   30:310–314. DOI: 10/gdkw3j.
473 Hatzipanagos S, Gregson J. 2014. The role of open access and open educational resources: A distance
474   learning perspective. *ICEL-2014 Hosted by The Federico Santa María Technical University Valparaiso
475   Chile*:265.
476 van Leeuwen TN, Tatum C, Wouters PF. 2018. Exploring possibilities to use bibliometric data to monitor
477   gold open access publishing at the national level. *Journal of the Association for Information Science and
478   Technology* 69:1161–1173. DOI: 10.1002/asi.24029.
479 van Leeuwen TN, Costas R, Robinson-Garcia N. 2019.Indicators of open access publishing in the CWTS
480   Leiden Ranking 2019. *Available at https://www.cwts.nl/blog?article=n-r2w2a4&title=indicators-of-
481   open-access-publishing-in-the-cwts-leiden-ranking-2019*
482 Martín-Martín A, Costas R, van Leeuwen TN, López-Cózar ED. 2018a. Unbundling Open Access
483   dimensions: a conceptual discussion to reduce terminology inconsistencies. DOI: 10.17605/osf.io/7b4aj.
484 Martín-Martín A, Costas R, van Leeuwen TN, Delgado López-Cózar E. 2018b. Evidence of open access
485   of scientific publications in Google Scholar: A large-scale analysis. *Journal of Informetrics* 12:819–841.
486   DOI: 10.1016/j.joi.2018.06.012.
487 Meneghini R, Mugnaini R, Packer AL. 2006. International versus national oriented Brazilian scientific
488   journals. A scientometric analysis based on SciELO and JCR-ISI databases. *Scientometrics* 69:529–538.
489   DOI: 10.1007/s11192-006-0168-z.
490 Minniti S, Santoro V, Belli S. 2018. Mapping the development of Open Access in Latin America and
491   Caribbean countries. An analysis of Web of Science Core Collection and SciELO Citation Index (2005–
492   2017). *Scientometrics* 117:1905–1930. DOI: 10/gfrb6b.

493   Moedas C. 2015. *Open Innovation, Open Science, Open to the World*. European Commission's
494   Directorate-General for Research & Innovation (RTD).
495   Packer AL. 2009. The SciELO Open Access: A Gold Way from the South. *Canadian Journal of Higher*
496   *Education* 39:111–126.
497   Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S.
498   2018. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles.
499   *PeerJ* 6:e4375. DOI: 10.7717/peerj.4375.
500   Robinson-Garcia N, Costas R, van Leeuwen TN. 2019. Indicators of Open Access for universities.
501   *arXiv:1906.03840 [cs]*.
502   Schmidt B, Kuchma I. 2012. *Implementing Open Access Mandates in Europe: OpenAIRE Study on the*
503   *Development of Open Access Repository Communities in Europe*. Universitätsverlag Göttingen.
504   Schomberg R von, Holbrook BJ, Oancea A, Kamerlin SCL, Ràfols I, Jacob M, Wouters P. 2019.
505   *Indicator frameworks for fostering open knowledge practices in science and scholarship*. Brussels:
506   European Commission's Directorate-General for Research & Innovation (RTD).
507   Torres-Salinas D, Robinson-Garcia N, Moed HF. 2019. Disentangling Gold Open Access. In: Glänzel W,
508   Moed HF, Schmoch U, Thelwall M eds. *Handbook of Science and Technology Indicators*. Cham:
509   Springer, 129–144.
510   Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, van Eck NJ, van Leeuwen TN, van
511   Raan AFJ, Visser MS, Wouters P. 2012. The Leiden ranking 2011/2012: Data collection, indicators, and
512   interpretation. *Journal of the American Society for Information Science and Technology* 63:2419–2432.
513   DOI: 10.1002/asi.22708.

514

Figure 1

Workflow followed to identify OA types based on Unpaywall data.

Source: van Leeuwen, Costas and Robinson-Garcia 2019
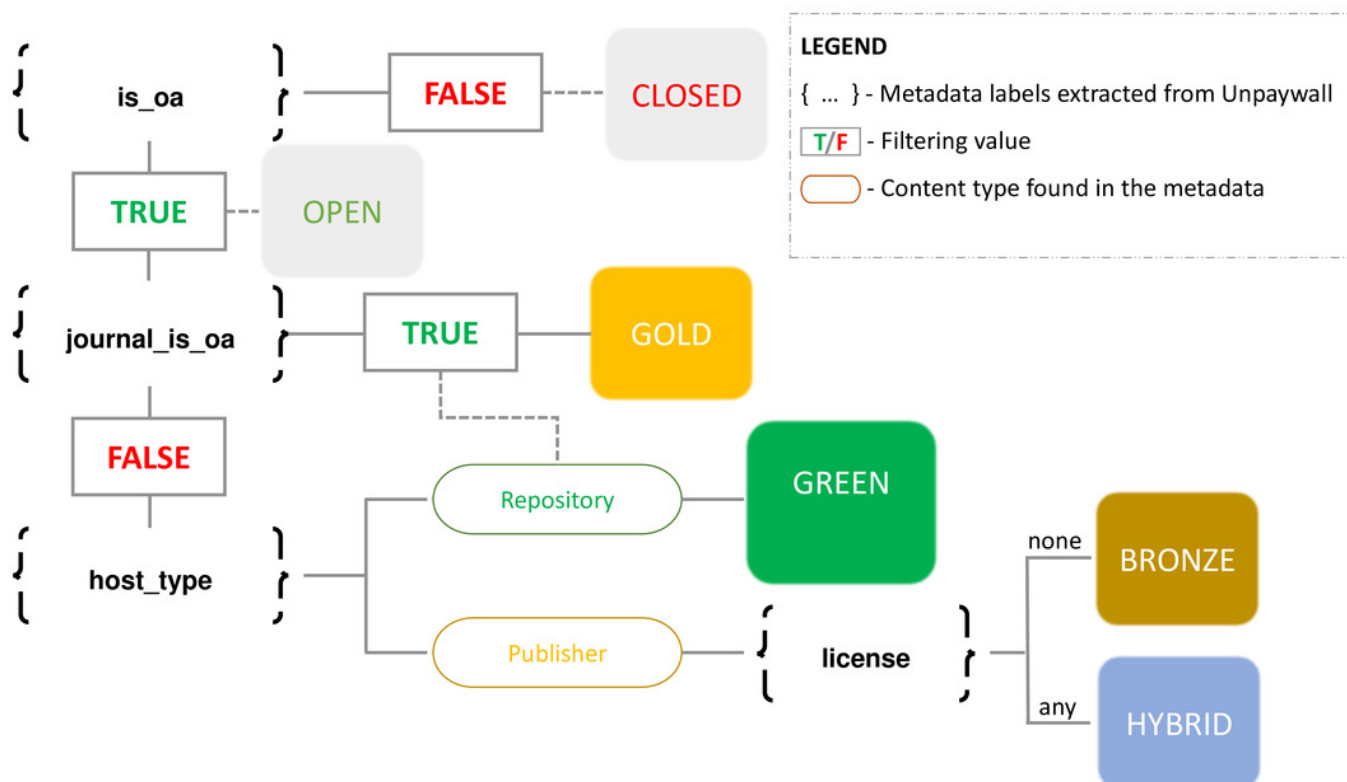
Figure 2

Total number of documents in open access by type and overlap.
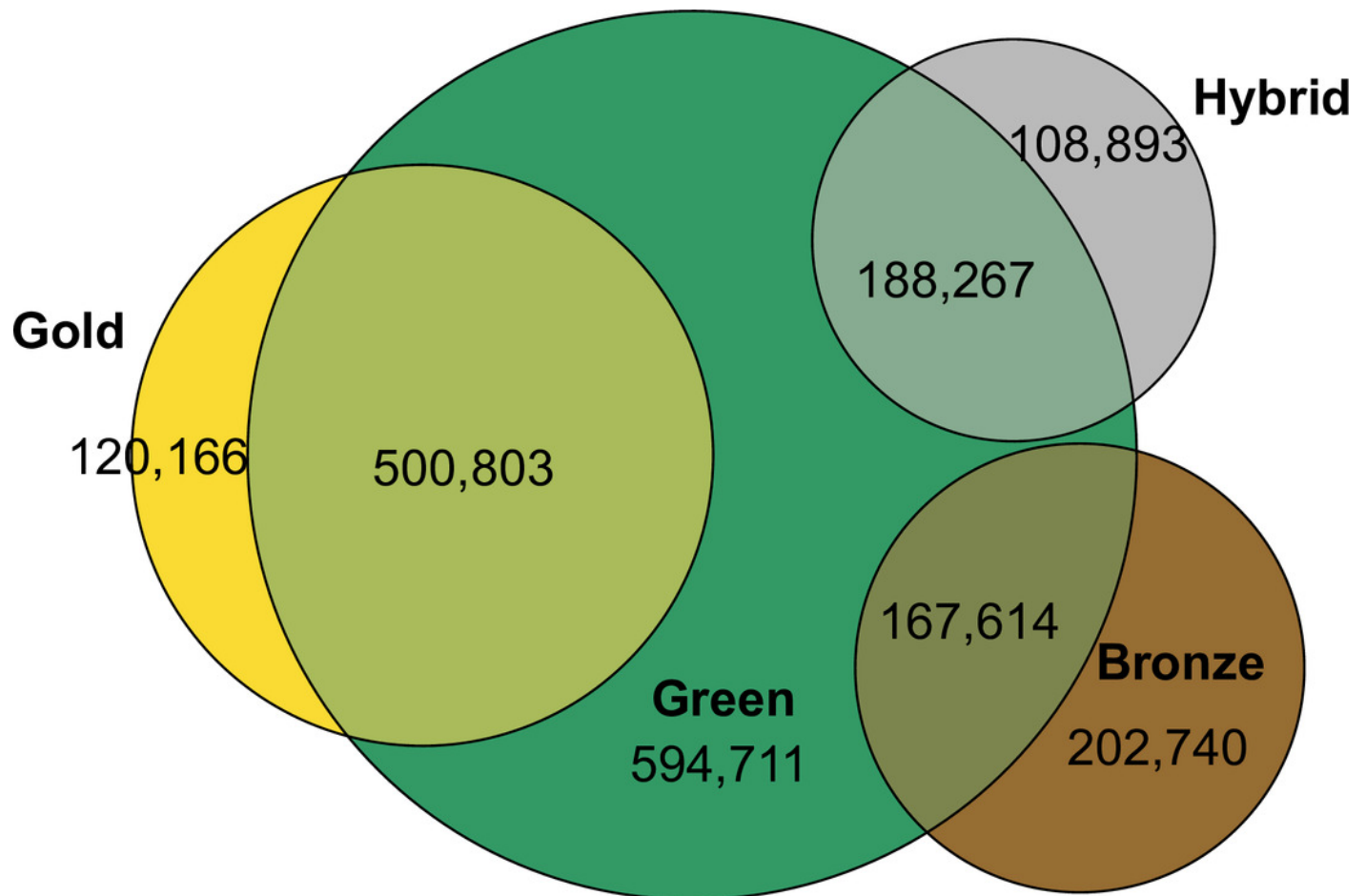
Figure 3

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.
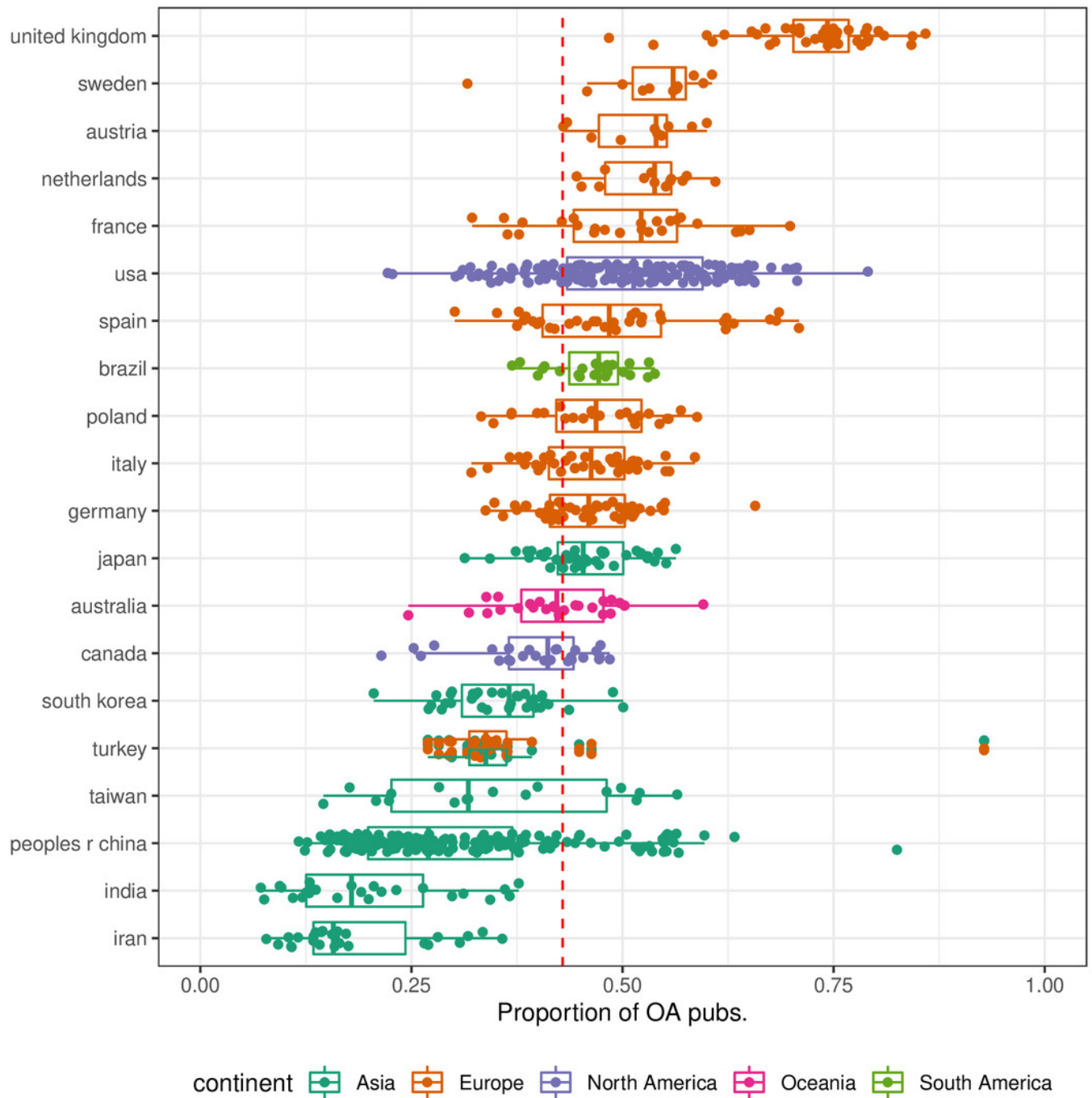
Figure 4

Proportion of OA publications of the set of universities analysed for A green OA, B gold OA, C hybrid OA and D bronze OA. Only countries with at least 10 universities included are shown. Countries are ordered based on the median value of the share of OA p
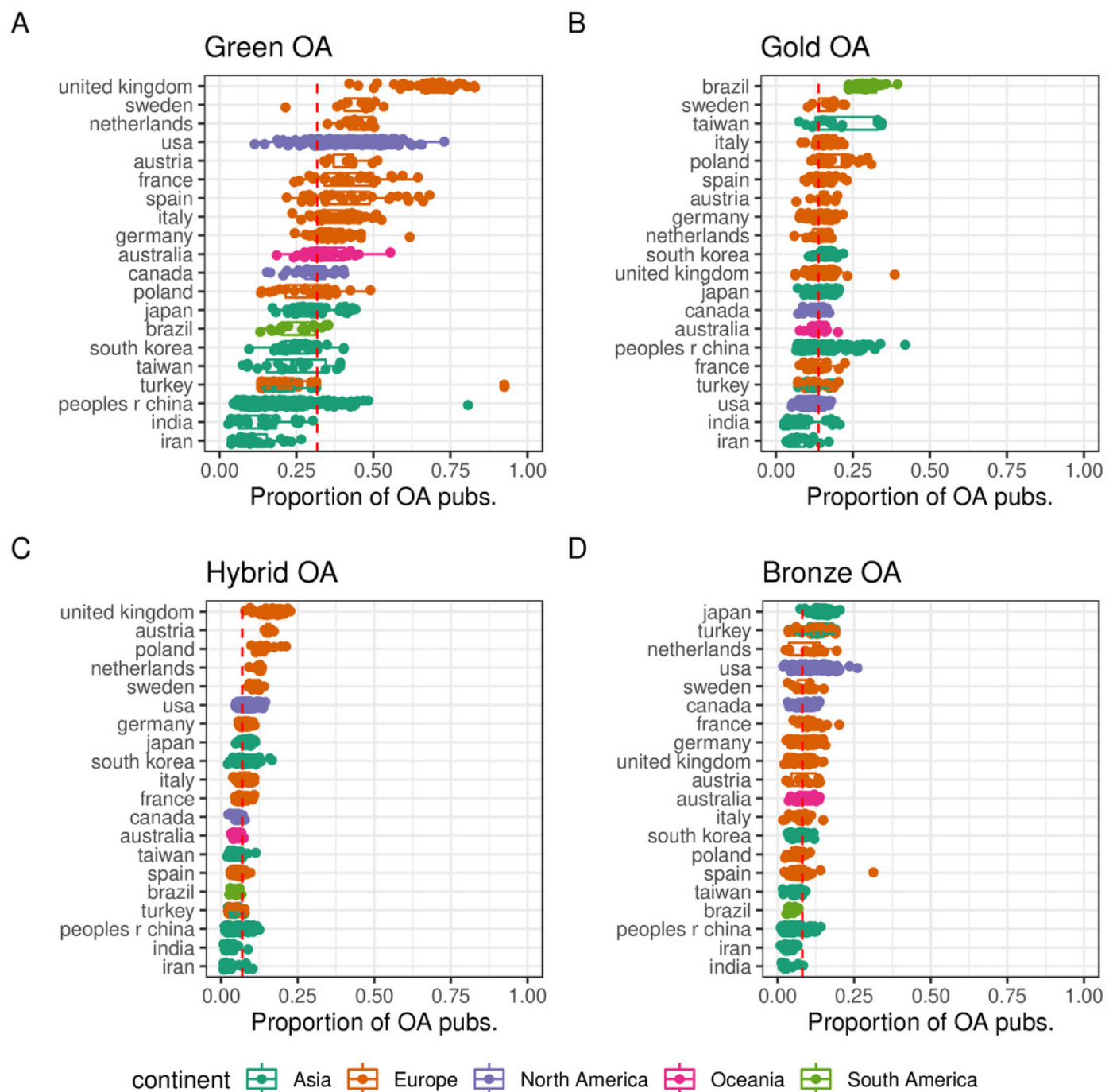
Figure 5

Proportion of OA publications of universities for each type of OA and for all OA types by field for universities worldwide.

Colors refer to each of the OA types. Orange: All OA; brown: bronze OA; yellow: gold OA; green: green OA; grey: hybrid OA.
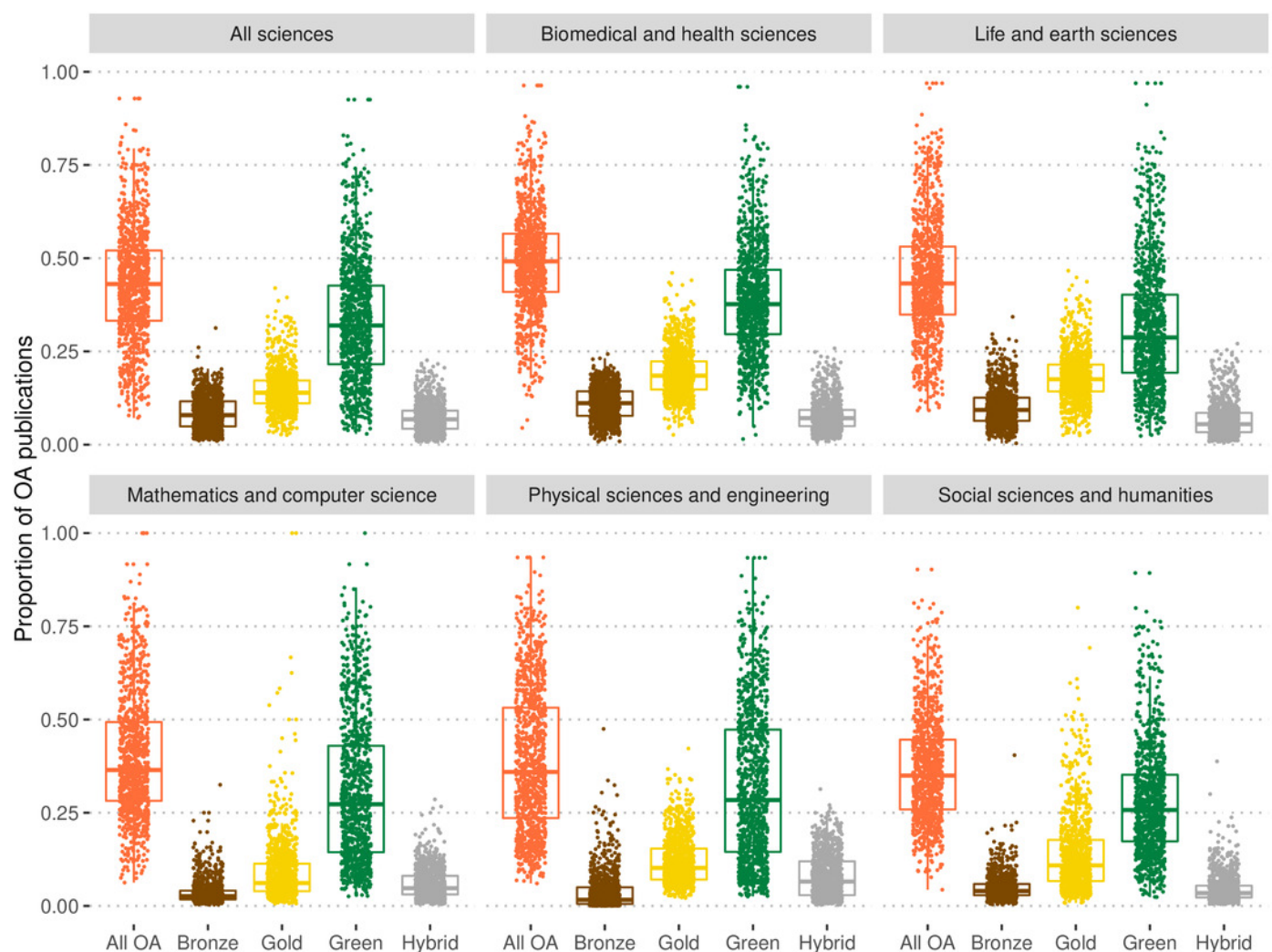
Figure 6

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

Colors refer to each of the OA types. Orange: All OA; brown: bronze OA; yellow: gold OA; green: green OA; grey: hybrid OA.
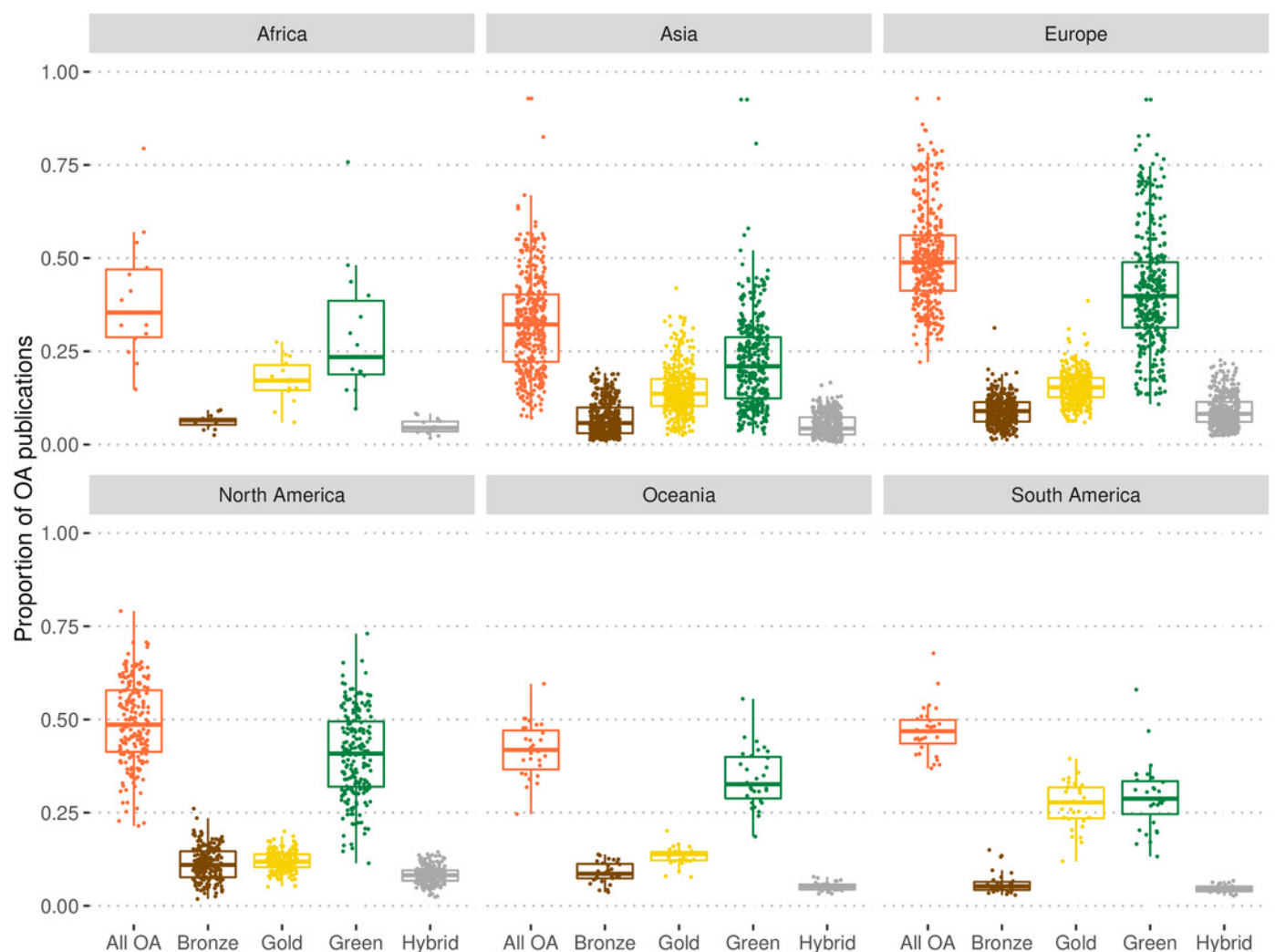
Figure 7

An example of OA disciplinary profiles for top 3 universities with the largest output for North America (A-C), Europe (D-F), Africa (G-I), South America (J-L) and Asia (M-O).

Colors refer to each of the OA types. Brown: bronze OA; yellow: gold OA; green: green OA; grey: hybrid OA.

A **Harvard Univ**

B **Univ Toronto**

C **Johns Hopkins Univ**

D **Univ Cambridge**

E **Univ Oxford**

F **Univ Coll London**

G **Univ Cape Town**

H **Univ Witwatersrand**

I **Cairo Univ**

J **Univ São Paulo**

K **Univ Estadual Paulista**

L **Univ Campinas**

M **Univ Chinese Acad Sci**

N **Shanghai Jiao Tong Univ**

O **Zhejiang Univ**
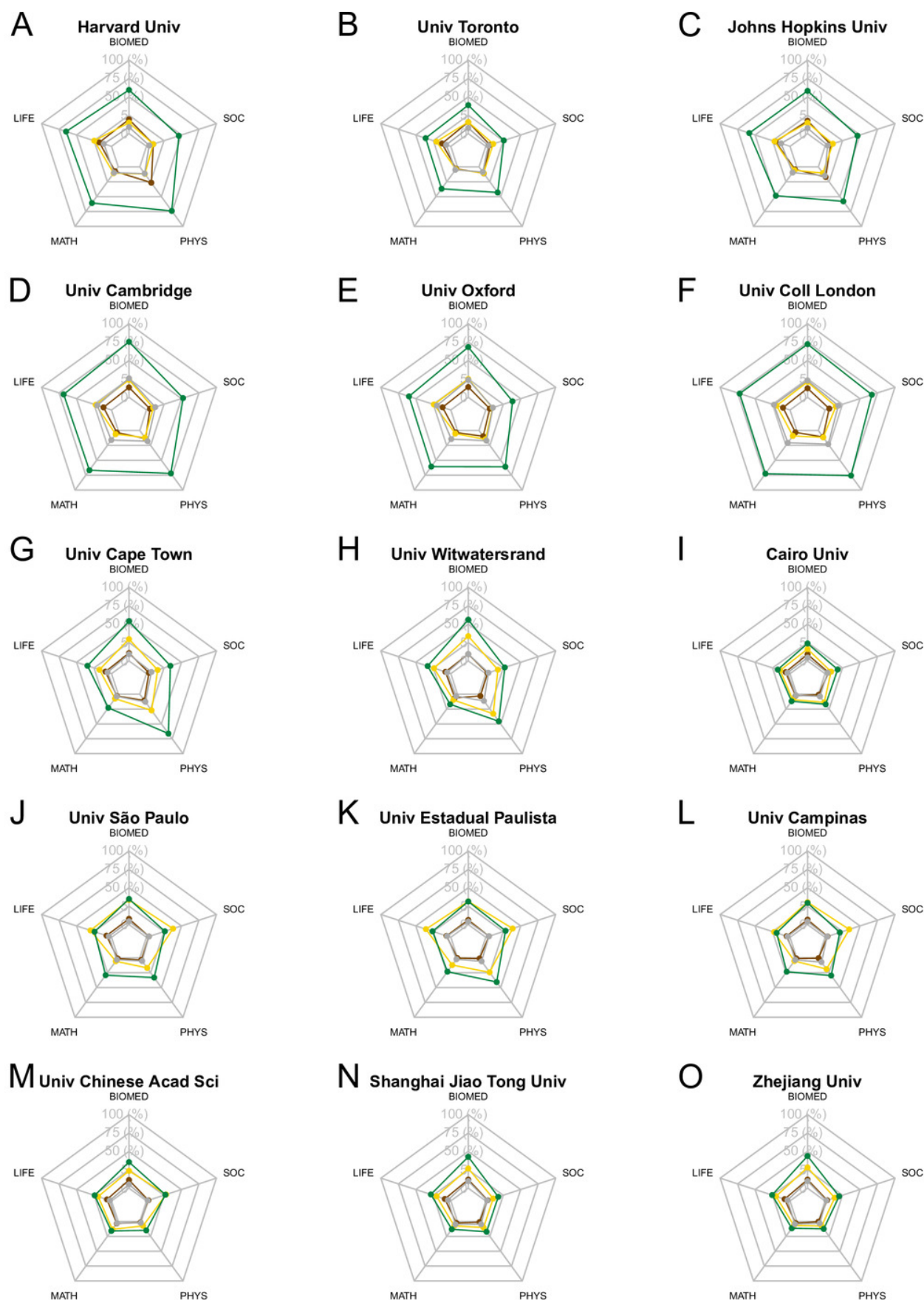
Figure 8

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
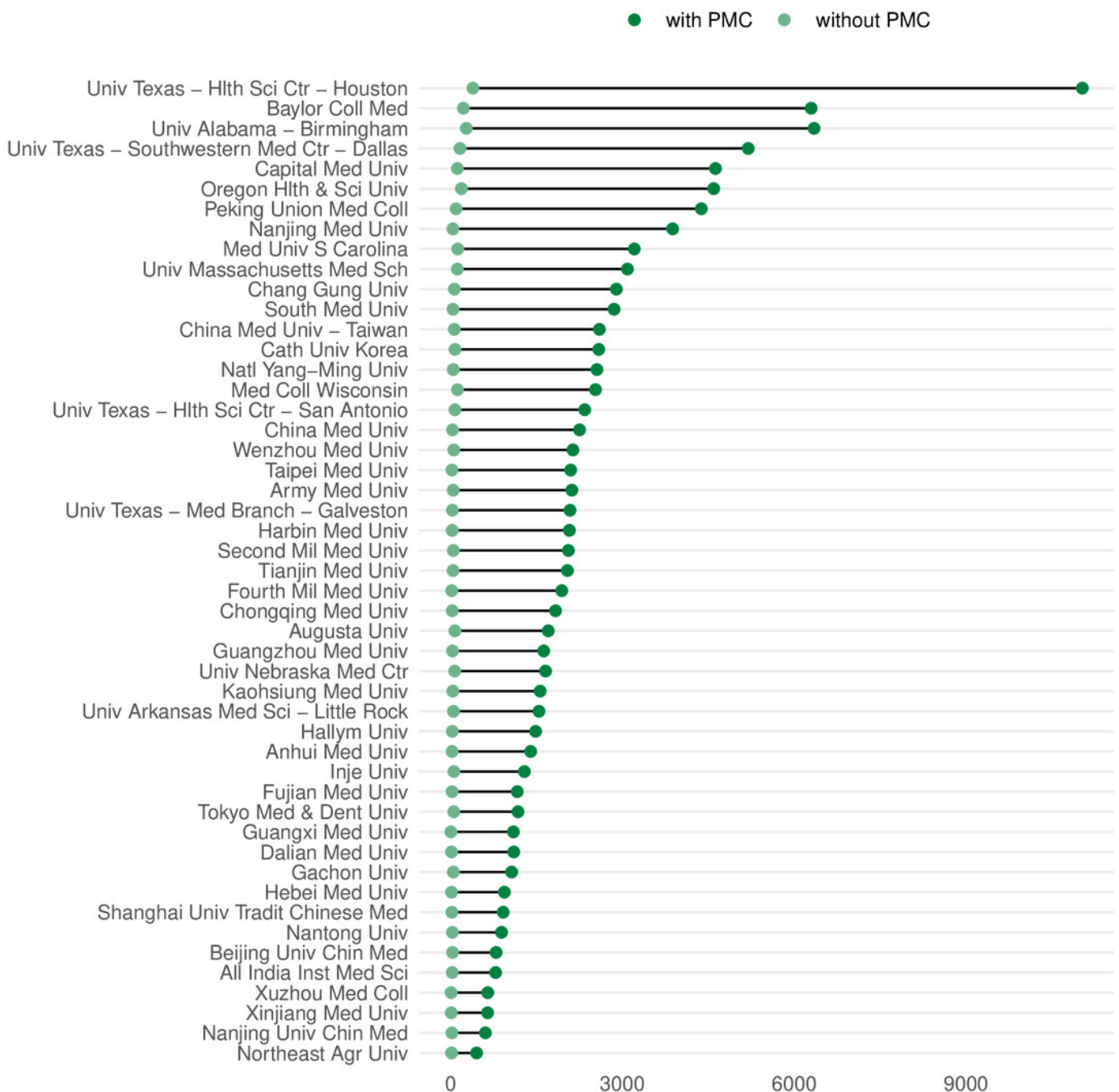
Figure 9

Share of Gold OA publications by country by type of publications.

Blue: Publications in national OA journals; Orange: Publications in APC OA journals; Green: 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.

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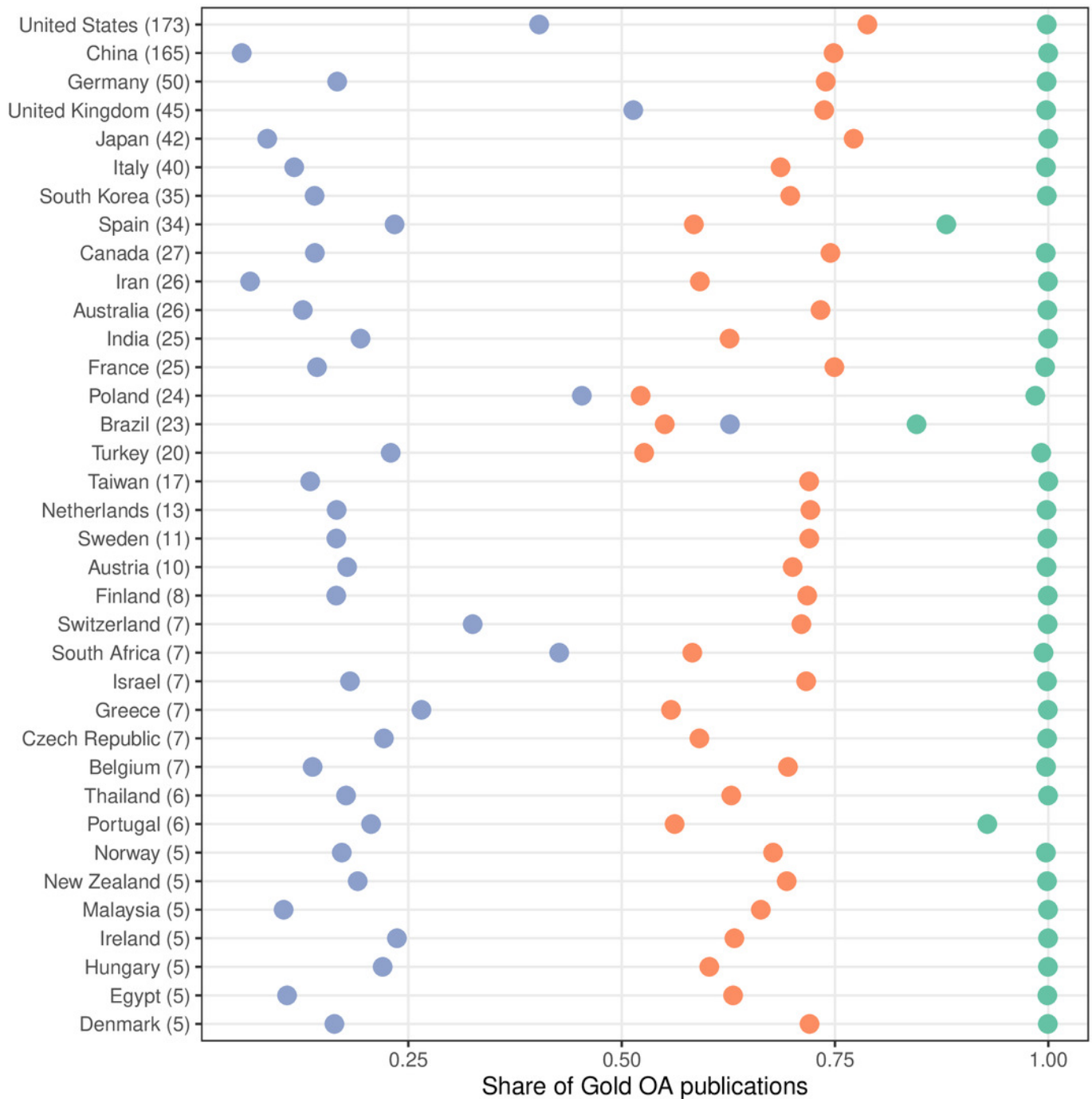
Table 1. Top 20 universities with the highest share of their output available through green OA.

URL strings used are available in the S1 Supplementary Material. *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.

University	Country	Pubs	Green pubs	% Green pubs in Repository*
Bilkent Univ	Turkey	2,008	1,858	97.7 – 97.7
City Univ London	United Kingdom	2,569	2,131	88.3 – 88.6
Durham Univ	United Kingdom	7,452	6,159	84.9 – 85.1
Hong Kong Polytech Univ	China	9,816	7,925	5.1 – 96.2
London Sch Hyg	United Kingdom	7,237	5,817	76.2 – 76.7
Univ Strathclyde	United Kingdom	4,847	3,830	88.9 – 89.0
Univ St Andrews	United Kingdom	5,780	4,497	79.2 – 79.7
Loughborough Univ	United Kingdom	4,274	3,271	85.9 – 86.2
Univ Pretoria	South Africa	6,432	4,873	93.7 – 93.7
Univ Leeds	United Kingdom	11,948	8,994	82.0 – 82.3
Univ Glasgow	United Kingdom	12,024	8,975	77.9 – 78.3
Univ Bath	United Kingdom	5,142	3,808	68.3 – 68.9
Univ Edinburgh	United Kingdom	18,139	13,415	55.2 – 58.2
Caltech	United States	13,481	9,834	69.2 – 69.4
Univ Bristol	United Kingdom	14,297	10,418	62.3 – 62.8
Univ Reading	United Kingdom	4,720	3,408	84.7 – 84.9
London Sch Econ	United Kingdom	3,525	2,534	79.4 – 79.83
Univ Coll London	United Kingdom	35,352	25,366	62.2 – 62.6
Univ Sussex	United Kingdom	5,510	3,931	69.1 – 69.7
Univ Warwick	United Kingdom	10,706	7,644	59.4 – 59.9

Top 20 countries with the highest share of distinct green OA publications coming from PMC.

Country	Green OA	PMC	PMC only	% Gold	% Bronze	% Hybrid
Taiwan	18,841	14,748	12,337	5.59	0.52	0.75
South Korea	43,425	34,066	26,995	13.27	1.98	2.67
China	190,201	138,931	114,228	67.32	8.66	14.46
Thailand	5,166	3,987	2,578	61.05	11.14	12.47
Lebanon	819	620	386	61.77	10.97	11.13
Egypt	3,604	2,394	1,617	63.53	9.61	11.53
Japan	59,787	34,289	24,104	58.30	17.58	14.41
Singapore	10,717	6,637	4,266	61.22	13.56	12.88
Malaysia	8,675	4,718	3,345	81.37	4.60	7.31
Poland	19,672	10,222	7,404	56.54	5.34	29.94
Pakistan	1,344	638	496	80.41	3.76	9.09
Austria	18,208	10,554	6,471	45.26	10.79	31.20
Canada	71,913	45,445	25,121	51.15	16.98	11.55
Iran	8,412	4,408	2,931	47.84	5.24	26.66
Brazil	35,134	18,901	11,707	76.16	7.40	6.09
India	10,475	4,923	3,414	67.13	8.61	8.27
USA	522,934	383,483	169,403	30.14	17.96	11.11
Israel	16,761	8,750	5,407	51.77	16.32	13.46
Mexico	6,133	2,758	1,924	71.86	10.30	6.82
Saudi Arabia	10,042	5,211	3,108	71.73	7.29	9.56