Scientometrics (2021) 126:2523–2556
https://doi.org/10.1007/s11192-020-03663-7

# Openness trends in Brazilian citation data: factors related to the use of DOIs

Rogério Mugnaini[1] · Grischa Fraumann[1,2] · Esteban F. Tuesta[1] · Abel L. Packer[3]

Received: 31 January 2020 / Published online: 17 August 2020
© Akadémiai Kiadó, Budapest, Hungary 2020

**Abstract**
Digital object identifiers (DOIs) are important metadata elements for indexing and interoperability, as well as for bibliometric studies in times of openness. This study analyses the use of DOIs in the cited references of articles by authors from Brazilian institutions, their possible influencing factors and differences among areas of knowledge. It measures the extent to which the citation datasets are open for reuse by others in terms of the availability of DOIs. 226,491 articles were retrieved from Web of Science (2012–2016), making a total of 8,707,120 cited references, 68% of which include DOIs. The results showed that the hard sciences have higher percentages of DOIs in their cited references. The factor type of collaboration showed higher percentages when there is international collaboration, being significantly different from the other categories. However, when the analysis was conducted inside the areas, the international collaboration was found to be different particularly in the soft sciences and a couple of other areas. The articles with DOI attributed, as well as those with mention of research funding, had a significantly higher percentage, even in the interaction with the areas of knowledge. Among the open access routes the green routes showed the highest percentages, followed by golden (DOAJ and other) and Bronze, but green routes articles proved to be not significantly different from those not openly accessible. Finally, the principal collaborating countries also showed the greatest influence on the DOI attribution, with the exception of Peru and South Africa. Our findings provide evidence that studies on the availability and usability of DOIs can assist researchers, by underlining the importance of making greater use of this persistent identifier, as well as to provide consistency to citation analysis.

**Keywords** Digital object identifier (DOI) · Citation data · Areas of knowledge · Openness of research · Research funding · Brazil

✉ Rogério Mugnaini
mugnaini@usp.br

Extended author information available on the last page of the article

## Introduction

Openness of research has brought about a new *modus operandi* of Science and Technology on a global scale. Several national, international and supranational open science initiatives are currently being carried out. This is leading to an improvement in classical text-based scientific communication, as research outputs become more freely available. Research data, for example, is being made directly available as well as being analysed in publications, and this means it can be reused and subjected to greater scrutiny (Albagli et al. 2014). Openness is a key factor in strengthening research and improving its availability and integrity.

An important driver of open science is to enhance the quality of all research outputs, so that they can serve as active agents within the global flow of scientific information. It was the inclusion of research data as an object of communication and source of citation that led to the drafting of the FAIR Guiding Principles (findable, accessible, interoperable, and re-usable) as a more comprehensive abstraction of object identity, location and interoperability of research outputs (Juty et al. 2020). A key feature of FAIR Principles is to endow each research object with a unique identifier and this converges with the fact that Digital Object Identifiers (DOIs) are already widely used and form apart of research outputs around the world (Wang et al. 2017).

A DOI is a persistent identifier (PID) assigned to a digital object—intended to be independent of URL (Uniform Resource Locator) changes—that was first released in 1997 (Boudry and Chartron 2017). “A DOI consists of a unique character string that identifies an entity in a digital environment” (Kraft et al. 2017), and a PID, in turn, is a “persistent, unique and globally resolvable identifier that is based on an openly specified schema” (Collins et al. 2018). The main purpose of creating and developing the DOI was to provide a unique identifier for any form of web-based published content and it has been promoted by the publishing industry. DOIs are assigned by several domain-centric registration agencies (International DOI Foundation 2019). Crossref is the most important DOI provider in scientific communication and was created by 6 of the major commercial publishers. In 2019, it had more than 11,000 members, most of them being small publishers and individual journals; there was a total of 109,440,179 registered DOIs, including 79,777,806 articles from 72,149 journal titles (Crossref 2020).

The presence of DOIs in cited references contributes to enhancing the transparency of the related data, as an additional part of the openness of research. As Jacsó (2004) pointed out, authors and publishers play a key role in this process, particularly with regard to the attribution of DOIs in the cited references. The author suggests that the current initiatives resulting from the use of DOI can help to solve an old problem—how to unify citation styles,—in a practical manner. Furthermore, this meant bibliometric studies could be undertaken that had previously only been possible at bibliographic databases of a few research centres in bibliometrics, which was very expensive (Costas 2017). In light of this, the bibliometric community also supports having a freely available list of references (Sugimoto et al. 2018) also a clear expression of openness of research.

The Initiative for Open Citations (I4OC) made a significant advance in the field of open access publishing, by urging publishers to make reference lists of their articles available in open access, so that, among other measures, the number of DOIs could be analysed. This resulted in 59% open references, out of 47.6 million articles, being made accessible via Crossref in September 2019. At the beginning of I4OC, only 1% of open references had been made available via Crossref. This amount of data is the result of an information flow that must be as inclusive as possible, and requires a cultural change in the concept of research, as Open Science has been advocating. Moreover, open initiatives in science should be implemented as a means of avoiding the mistakes of the past, and overcoming barriers—mainly linguistic ones (Gibbs 1995; Meneghini and Packer 2007; Mugnaini et al. 2019). In addition, to be part of this movement, changes must be made by each of its actors.

On the one hand, it should be noted that not all journals provide DOIs for their articles, as pointed out by Rodríguez-Yunta (2013), with regard to the "academic silence around the DOI in Spain". Conversely, the author mentions that the DOI was approved as a standard (ISO 26324) in 2012, and that this could have been an important factor in stimulating its use. In Brazil, ABNT/NBR 6023 is the official standard for citation styles—from the Brazilian Technical Standards Association. It was updated as recently as 2018, when it replaced the earlier version from 2002 that did not mention DOI in its guidelines. Fortunately, in Brazil there are benefits for publishers, such as the indexing criteria of the SciELO Project, which requires adopting formal procedures and following certain publishing guidelines (Montanari and Packer 2014). Not surprisingly, Martín (2013) noted that SciELO Brazil was the only collection where all the journal editors had begun making use of the DOI attribution at that time. Nevertheless, as Jacsó (2004) argued, authors play a crucial role as well as editors.

## Usefulness and presence of DOIs in the scientific literature

The analysis of DOIs in certain information sources has attracted a good deal of attention among scholars. The use of DOIs as a key feature element of cited references is a means of enhancing the integrity of research, together with its dissemination, access, interoperability and follow-up studies including the assessment of its scientific impact through citation metrics. DOIs or other persistent identifiers are also required for altmetric studies (Zahedi et al. 2014), as altmetric events represent another dimension of openness; this is best traced through publications with a persistent identifier, which in most cases is a DOI (Haustein 2016). DOIs may assist in allowing the reuse and interconnectivity of research results, and are also useful for the retrieval of open access articles from various data sources (Piwowar et al. 2018). Ultimately, research objects without DOI will increasingly disappear from the global flow of digital scientific information. That is why Gorraiz et al. (2016) encourage researchers to favour journals with established DOI assignment when deciding where to submit manuscripts.

Bibliographic databases are currently an important channel for the dissemination of research results, since, as well as being employed for bibliometric studies, they can be used for carrying out literature reviews by researchers, students and other interested stakeholders (Gusenbauer 2019). Many other initiatives exemplify the usefulness of DOIs for the following reasons: (a) their multiple database coverage, that is regarded as one of the quality criteria for metadata (Hug and Brändle 2017); (b) their presence in databases such as Web of Science (WoS) and Scopus, that are recommended as one of the most effective means of tracing scientific output (Wouters et al. 2019); (c) they can be used to calculate the percentage of open access publications in WoS (Piwowar et al. 2018); and (d) they can be used to compare publications from different databases, through citation matching algorithms (Hug and Brändle 2017; van Eck et al. 2018).

DOIs can also be used to match publications in large CV databases, such as the Brazilian CV Lattes Platform. Rubim and Braganholo (2017) found that the number of DOIs in the CV Lattes Platform is too low for a comprehensive analysis (56.78% of the journal papers and 91.98% of the conference papers do not have a DOI). In addition, there are statistics made available by Crossref on how many requests for DOIs have been made by certain countries, for example in the case of Brazil (Ruiz 2011). The number and origin of DOIs have also been analysed in other online platforms, such as language-specific Wikipedia sites (Kikkawa et al. 2016).

Gorraiz et al. (2016) analysed percentage of publications with DOIs from 2005 to 2014 in WoS and Scopus, considering the whole database. Their study revealed significant differences between academic disciplines or areas of knowledge, especially with regard to citable documents (articles, proceeding papers and reviews) in WoS: Science Citation Index (SCI) and Social Science Citation Index (SSCI) peaked around 90% in 2014, while Arts & Humanities Science Citation Index (A&HCI), surpassed 50% in 2013. In Scopus the same type of documents also presented around 90% in hard sciences in 2014; while Social Sciences & Humanities, that were taken together, reached 80%. Haustein et al. (2015) also used the whole WoS database, restricted to articles from 2012, and with a slightly different aggregation field. The following percentages were found: Biomedical and Health Sciences, Mathematics and Computer Science and Natural Sciences and Engineering (~80%); Life and Earth Sciences (~74%); and Social Sciences and Humanities (~63%). As can be seen, the different databases, aggregation of areas and periods show variations, since the DOIs were assigned in many areas and throughout the time.

Gumpenberger et al. (2016) argue in favour of the online presence of articles in social sciences and humanities, that could be improved by the use of DOIs in three universities studied (University of Vienna, Humboldt University of Berlin and University of Zurich), which presented ~65% of publications with DOIs in SSCI, compared with ~26% in A&HCI. Non-English publications is another factor which is an obstacle to international visibility. According to the authors, this problem can be overcome by providing metadata information in English in international databases.

The percentage of DOIs attributed to research results from certain regions and countries in the world, is another factor that has been studied in bibliometrics, for example in the case of African Journals Online (AJOL), a major bibliographic database for Africa (Fasae and Oriogu 2018). Martín (2013) analysed the percentage of journals from the collections of different SciELO Ibero-American countries that assign DOIs to articles, and concluded that less than 50% of the countries did so at that time. Although it is a country in which all its journals are assigning DOIs in the present, Brazil (as well as other countries) has been striving to assign the DOI to the entire collection retrospectively.

Other scholars have studied a dataset of 496,665 articles from PubMed (1966 until 2015), 40.48% of which contained a DOI, but if only the year 2015 was taken into account, the number of DOIs rose to 86.42% (Boudry and Chartron 2017). They found that, while editors from some countries gradually implemented DOIs during this period (the United States, United Kingdom and Netherlands), there were no DOIs available to others (Russia, the Czech Republic and Romania). In the same way, Wang et al. (2017) reported a relatively low use of DOIs in China compared with other countries. This shows that at certain times, new practices of scientific communication are not being adopted, nor are they taking place in different places with the same degree of intensity.

## Research questions and hypotheses

Although Open Science is a global phenomenon, it should be recognised that effective changes in scientific culture do not take place in an equitable way, with regard to different countries and areas of knowledge. In the case of Brazil, is important to consider its continental dimensions—like India, China and Russia—but it is the most active open access publishing—while the others are in a contrasting situation (Wang et al. 2018). However, despite open access being an important initiative under the vast umbrella of open science, many other challenges need to be faced (Albagli et al. 2014). This makes Brazil an interesting case to study, besides the fact that its scientific output currently involves a low level of international collaboration, which can be explained by its geographic and multidisciplinary diversity (Digiampietri et al 2019).

As previous studies largely relied on the presence of DOIs in articles (as source data), little is known about the use of DOIs in cited references, and much less about factors regarding their use. This phenomenon needs to be investigated in the future since in many cases authors may have failed to provide information about the DOI in the bibliographic references, even though it exists. This gives rise to three research questions (RQ) and, when appropriate, the derived hypotheses (H) are defined:

(RQ1) To what extent do different factors influence the use of DOIs in the cited references of articles when they include at least one author from a Brazilian institution?

**(H1)** Since there are clearly different percentages of DOIs of articles from different areas and countries, and new practices are not being adopted at the same time, soft sciences and articles written without international collaboration, have a lower number of DOIs in the cited references.

(RQ2) What patterns of influence from other factors can be found among areas of knowledge and open access routes?

**(H2)** Areas where there is less international collaboration tend to be more influenced by the presence of a foreign author, showing greater presence of DOIs in the cited references of articles written with international collaboration.

(RQ3) Which countries collaborate most and/or exert the greatest influence on the DOI attribution in cited references?

## Method

### Data

The data covers 226,491 articles that contain at least one author from a Brazilian institution, published in journals indexed in WoS from 2012 to 2016. The manual download (500 registers each time) was done from Web of Science (from Clarivate Analytics) in December 2018, using the following search limits:

Author Address: (CU = BRAZIL OR CU = BRASIL)

Year Published: (PY = 2012 OR PY = 2013 OR PY = 2014 OR PY = 2015 OR PY = 2016)

Document Type: (DT = Article OR DT = Proceedings Paper OR DT = Review)

Web of Science Core Collection Citation Indexes: SCI-EXPANDED, SSCI, A&HCI

We retrieved the full register, including the cited references of the articles that totalized 8,707,120. We included only citations to journal articles, discarding those related to dissertations and other, that can be found together.

### Variables

The article was considered as unity of analysis, being the percentage of DOIs in its cited references (when it was related to journal articles) the dependent variable. Among the set of cited references 5,923,823 (68%) present a DOI. Although, it was not our goal to check for correctness, for example, testing the access to the source article (Piwowar et al. 2018). After processing the information in the "Cited References Field" (CR)—separating the different cited references, discarding non journal articles and separating each piece of information inside the cited reference—we searched for the DOI information. Hence, we calculated the percentage of cited references with DOIs of each article.

The DOI is a relatively new information, and its usage is not established in certain research communities (Boudry and Chartron 2017; Martín 2013; Rodríguez-Yunta 2013). This leads to the observation that the oldest articles have the lowest percentage of DOI attributed by journals. We also understand that, when analysing the usage of DOIs in cited references, the absence of the information can be a consequence of the non-existence of DOI in the source, which could result in false negatives. It is worth considering also that this problem could mask the results when dealing with older cited references, even if several journals have made efforts to attribute DOIs retrospectively.

Table 1 shows the distribution of the percentage of DOIs in cited references by citing and cited years (showing cited years until 1993). The cells show color scales with higher values in blue, permitting the observation that the higher percentages concentrate in articles from 2014 and 2013 and cited references from 2010 to 2003.

In order to minimize the effect of absence of DOIs in older articles we defined 2002 as a limit (guaranteeing a minimum of 60.2% in the upper area), resulting in two forms of calculation of the dependent variable, that is the percentage of DOIs in the cited references of the articles. As such, we considered the following criteria:

(1) the comprehensive set (all cited references), and;

(2) a restricted set of cited references (2002 onwards).

The independent variables are the factors we defined to test influence in the use of DOIs in cited references. They consisted of some characteristics of the citing articles, considering information of various bibliographic fields presented in WoS:

1. *Area of knowledge of the journal* we adapted the Essential Science Indicators™[1] (also from Clarivate Analytics), that proposes 22 broad areas, covering journals indexed in

[1] Retrieved from https://help.incites.clarivate.com/incitesLiveESI/ESIGroup/overviewESI/esiJournalsList.html.

**Table 1** Distribution of percentage of DOIs in references according to articles (2016–2012) and cited references (2016–1993) years

| Cited Years | Citing Years 2016 | 2015 | 2014 | 2013 | 2012 |
|---|---|---|---|---|---|
| 2016 | 60.2 | | | | |
| 2015 | 63.9 | 61.3 | | | |
| 2014 | 64.4 | 63.0 | 61.4 | | |
| 2013 | 64.5 | 64.0 | 64.7 | 62.1 | |
| 2012 | 65.1 | 64.7 | 66.1 | 65.4 | 61.0 |
| 2011 | 66.2 | 65.8 | 67.9 | 68.1 | 65.3 |
| 2010 | 67.5 | 67.1 | 69.9 | 70.1 | 68.1 |
| 2009 | 67.8 | 67.0 | 69.3 | 69.8 | 67.8 |
| 2008 | 67.5 | 67.4 | 69.8 | 70.1 | 68.0 |
| 2007 | 67.7 | 66.8 | 69.4 | 69.9 | 67.9 |
| 2006 | 66.9 | 66.4 | 69.3 | 69.3 | 67.4 |
| 2005 | 66.7 | 66.4 | 68.9 | 69.8 | 67.3 |
| 2004 | 67.7 | 66.6 | 69.9 | 70.1 | 68.1 |
| 2003 | 65.9 | 65.3 | 68.9 | 68.6 | 66.7 |
| 2002 | 64.1 | 62.2 | 65.7 | 65.8 | 63.3 |
| 2001 | 62.0 | 59.2 | 62.3 | 62.1 | 61.1 |
| 2000 | 60.8 | 57.2 | 60.3 | 60.7 | 58.7 |
| 1999 | 58.4 | 55.0 | 57.2 | 57.4 | 56.1 |
| 1998 | 57.4 | 52.4 | 55.4 | 55.8 | 54.3 |
| 1997 | 57.0 | 51.6 | 54.0 | 54.4 | 52.7 |
| 1996 | 56.1 | 49.5 | 51.2 | 51.7 | 51.0 |
| 1995 | 54.3 | 47.7 | 49.6 | 49.2 | 48.2 |
| 1994 | 52.4 | 45.4 | 47.4 | 47.8 | 45.5 |
| 1993 | 54.4 | 47.9 | 49.5 | 48.7 | 47.9 |

the Web of Science Core Collection (Science Citation Index Expanded and Social Sciences Citation Index only), to which we added the Human Science area (journals from Arts & Humanities Citation Index), resulting in 23 categories. We used the "ISSN" (SN) field to match journal and the respective broad area (this classification associates each journal to just one area);

2. *Type of collaboration* based in the "Author" (AU) and "Author Address" (C1) fields, the following criteria were applied: if the article has at least one foreign author in the collaboration it was *international*, else; if there were authors from more than one institution, *national*, else; if there was more than one author, *institutional*, else; *single author*;
3. *Status of the article (closed access/open access route)* the field "Open Access Indicator" (OA)[2] that identifies if the article is openly accessible or not. If the article is in open access it was classified as Gold (Directory of Open Access Journals—*DOAJ* or *Other*), Bronze or Green (Published or Accepted). To these categories we added "Gold SciELO"

[2] "Open access status is provided across the Web of Science platform as a result of a partnership with Impactstory, a not-for-profit organization that recently launched a knowledgebase of Open Access (OA) content. This knowledge base makes it possible to discover and link to legal Gold or Bronze (free content at a publisher's website) and Green (e.g., author self-archived in a repository) OA versions. This partnership improves discoverability and access to article-level OA versions not only by adding more links to OA content, but also by prioritizing the links to the best version of OA content when multiple versions of an article are available". Retrieved from https://images.webofknowledge.com/images/help/WOS/hp_results.html#dsy10670-TRS_open_access.

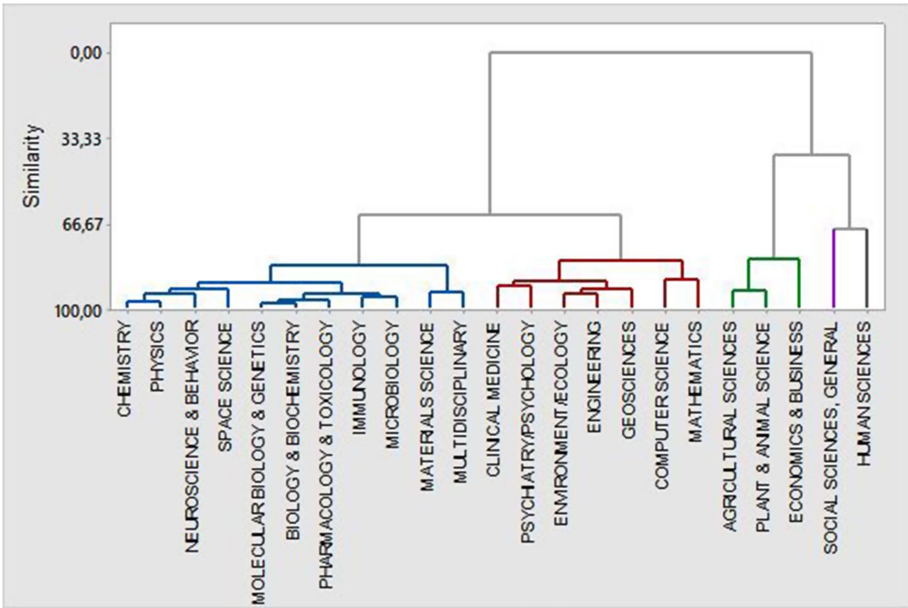


**Fig. 1** Dendrogram of areas of knowledge, considering different statistics of the percentage of DOIs in the cited references of the articles (all and from 2002 onwards)

when the journal was also indexed in Scientific Electronic Library Online (SciELO), which we detected via the "ISSN" (SN) field;

4. *Presence of DOI in the article* information presented in the "Digital Object Identifier (DOI)" (DI) field;
5. *Research funding* we checked if there was mention of research funding in "Funding Agency and Grant Number" (FU) field;
6. *Foreign co-author affiliation country* using the "Author Address" (C1) field, each foreign country was identified.

## Data analysis

First of all, we conducted a descriptive statistical analysis, to explore how the dependent variables behave with regard to each category/factor. After examining the areas of knowledge, we decided to group them on the basis of similar behavior patterns, since a part of them are very similar. Thus, we carried out an hierarchical cluster analysis with the following statistics of the first dependent variable (using the comprehensive set of cited references): first and third quartiles, median, mean, standard deviation and asymmetry coefficient of the percentage of DOIs in all the cited references of the articles. We also took note of other measures, that resulted from the variation caused by being restricted to the recently cited references (2002 onwards), that is to say, the difference of each of the statistics, with regard to the restricted and comprehensive sets [for example, to the mean (2002 onwards)—mean (all references)]. In the dendrogram of Fig. 1 we show five groups that displayed enough different profiles to enhance the analysis.

**Table 2** Distribution of the articles in the areas of knowledge, combined with the other factors—2012–2016

| Cluster | Area | Total | Type of collaboration | | | | DOI in the article? | | Mention of funding? | | Open acces route | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Single author | Institu-tional | National | Interna-tional | No | Yes | No | Yes | SciELO Gold | DOAJ Gold | Other Gold | Bronze | Green Publish. | Green Accept. | Not open |
| 1 | HUMAN SCIENCES | **4,695** | 2,755 | 626 | 1,027 | 287 | 2,567 | 2,128 | 4,658 | 37 | 590 | 109 | 206 | 9 | 2 | | 3,779 |
| 2 | SOCIAL SCIENCES, GENERAL | **16,536** | 2,270 | 4,732 | 6,667 | 2,867 | 4,502 | 12,034 | 13,850 | 2,686 | 6,462 | 444 | 144 | 287 | 129 | 83 | 8,987 |
| 3 | AGRICULTURAL SCIENCES | **21,681** | 96 | 6,034 | 12,497 | 3,054 | 5,027 | 16,654 | 12,106 | 9,575 | 10,679 | 2,359 | 68 | 1,098 | 182 | 16 | 7,279 |
| | ECONOMICS & BUSINESS | **2,771** | 205 | 811 | 1,060 | 695 | 821 | 1,950 | 2,423 | 348 | 283 | 222 | 12 | 16 | 15 | | 2,223 |
| | PLANT & ANIMAL SCIENCE | **25,852** | 445 | 6,673 | 11,923 | 6,811 | 6,235 | 19,617 | 8,480 | 17,372 | 7,142 | 1,377 | 261 | 1,850 | 412 | 29 | 14,781 |
| 4 | CLINICAL MEDICINE | **40,556** | 278 | 12,479 | 16,462 | 11,337 | 3,666 | 36,890 | 22,286 | 18,270 | 13,442 | 2,848 | 679 | 2,754 | 1,495 | 742 | 18,596 |
| | COMPUTER SCIENCE | **3,307** | 86 | 881 | 912 | 1,428 | 229 | 3,078 | 985 | 2,322 | 1 | 143 | 24 | 132 | 88 | 8 | 2,911 |
| | ENGINEERING | **10,998** | 265 | 3,466 | 3,719 | 3,548 | 778 | 10,220 | 3,860 | 7,138 | 716 | 317 | 73 | 252 | 83 | 9 | 9,548 |
| | ENVIRONMENT/ECOLOGY | **8,204** | 92 | 1,800 | 3,177 | 3,135 | 704 | 7,500 | 1,856 | 6,348 | 1,133 | 469 | 176 | 316 | 107 | 18 | 5,985 |
| | GEOSCIENCES | **4,332** | 63 | 804 | 1,264 | 2,201 | 232 | 4,100 | 984 | 3,348 | 525 | 376 | 69 | 278 | 96 | 3 | 2,985 |
| | MATHEMATICS | **4,918** | 663 | 754 | 1,244 | 2,257 | 641 | 4,277 | 1,029 | 3,889 | 130 | 159 | 60 | 562 | 27 | 11 | 3,969 |
| | PSYCHIATRY/PSYCHOLOGY | **3,292** | 142 | 720 | 1,021 | 1,409 | 482 | 2,810 | 1,817 | 1,475 | 903 | 212 | 29 | 81 | 75 | 126 | 1,866 |
| 5 | BIOLOGY & BIOCHEMISTRY | **9,336** | 75 | 2,474 | 3,774 | 3,013 | 465 | 8,871 | 1,851 | 7,485 | 1,366 | 1,078 | 129 | 1,169 | 550 | 185 | 4,859 |
| | CHEMISTRY | **16,363** | 149 | 5,064 | 6,030 | 5,120 | 1,326 | 15,037 | 2,910 | 13,453 | 2,509 | 886 | 151 | 686 | 235 | 69 | 11,827 |
| | IMMUNOLOGY | **4,091** | 12 | 807 | 1,524 | 1,748 | 178 | 3,913 | 891 | 3,200 | 294 | 1,201 | 164 | 366 | 353 | 179 | 1,534 |
| | MATERIALS SCIENCE | **6,480** | 62 | 1,649 | 2,474 | 2,295 | 283 | 6,197 | 1,496 | 4,984 | 1,060 | 367 | 38 | 149 | 67 | 26 | 4,773 |
| | MICROBIOLOGY | **4,160** | 15 | 880 | 1,732 | 1,533 | 375 | 3,785 | 616 | 3,544 | 474 | 872 | 96 | 473 | 276 | 80 | 1,889 |
| | MOLECULAR BIOLOGY & GENETICS | **5,128** | 25 | 1,044 | 2,143 | 1,916 | 119 | 5,009 | 652 | 4,476 | 279 | 884 | 204 | 1,370 | 313 | 168 | 1,910 |
| | MULTIDISCIPLINARY | **6,933** | 83 | 1,432 | 2,462 | 2,956 | 408 | 6,525 | 1,368 | 5,565 | 717 | 5,270 | 125 | 8 | 215 | 36 | 562 |
| | NEUROSCIENCE & BEHAVIOR | **5,796** | 54 | 1,618 | 1,985 | 2,139 | 90 | 5,706 | 1,547 | 4,249 | 673 | 419 | 146 | 597 | 224 | 244 | 3,493 |
| | PHARMACOLOGY & TOXICOLOGY | **6,146** | 31 | 1,740 | 2,717 | 1,658 | 863 | 5,283 | 1,193 | 4,953 | 682 | 370 | 48 | 720 | 396 | 83 | 3,847 |
| | PHYSICS | **12,794** | 686 | 2,201 | 2,770 | 7,137 | 159 | 12,635 | 1,525 | 11,269 | 346 | 1,415 | 898 | 710 | 731 | 7 | 8,687 |
| | SPACE SCIENCE | **2,122** | 59 | 199 | 176 | 1,688 | 26 | 2,096 | 174 | 1,948 | 12 | 9 | 38 | 985 | 227 | | 851 |
| **Total** | | **226,491** | **8,611** | **58,888** | **88,760** | **70,232** | **30,176** | **196,315** | **88,557** | **137,934** | **50,418** | **21,806** | **3,838** | **14,868** | **6,298** | **2,122** | **127,141** |
| *%* | | ***100%*** | ***4%*** | ***26%*** | ***39%*** | ***31%*** | ***13%*** | ***87%*** | ***39%*** | ***61%*** | ***22%*** | ***10%*** | ***2%*** | ***7%*** | ***3%*** | ***1%*** | ***56%*** |

Table 2 shows the distribution of articles in the clusters of areas, as well as their cross-distribution with the other factors (the independent variables listed above). The coloured scales illustrate how the cross-distribution between areas of knowledge and each of the factors are distributed. The darker blue cells highlight the concentration of Brazilian production in Clinical Medicine, Plant & Animal Sciences, Agricultural Sciences, Social Sciences and Chemistry, with more than 15 thousand articles on each, and making up 53% of the total. The Table also highlight the concentration of output: in *national collaboration* and *SciELO Gold* (Clinical Medicine, Plant & Animal Science, Chemistry, Physics and Agricultural Sciences); and not openly accessible (all areas).

Another observation concerns the categories of some factors with low frequencies (red cells in Table 2): single authorship in Clusters 3 to 5 while in Cluster 1 (Human Science) is the highest; articles without DOI and without mention to funding are scarce in Clusters 4 and 5 (hard sciences, with exception of Health Sciences); and among open access routes, the lowest frequencies are *Green Accepted* and *Other Gold* (practically in all areas) and *Green Published* (mainly in Clusters 1–4). The analysis of these interactions between factors must be treated cautiously, as the percentage of DOIs in these articles cited references, tends not to be robust.

To study the influence of the factors listed above we performed a General Linear Models (GLM) analysis. The explanation of the variation was explored through GLM because it allows a generalization of the distribution of errors, as well as the use of a link function that relates the mean of the dependent variable to the linear combination of the categories of the independent variables—which will result in a closer examination of the factors of influence. Once the dependent variable has binomial distribution, the logit transformation was used as link function. The independent variables listed above were considered as well as the interactions between the areas of knowledge against the other factors; and the open access routes against the other factors.

We built a Country *versus* Area matrix, in which the rows are the countries (affiliation of foreign authors) and the columns are the areas (the 23 areas of knowledge from Table 2), to analyze the international activity and influence on Brazilian research outputs. The values corresponding to each element of this matrix are the number of co-publications. We then used the measure RCA (Revealed Comparative Advantage) measure, as defined in Guevara et al. (2016), which is the ratio of the percentage of joint publications by area and country to the percentage of production by area and total. When analyzing the diversity of the influence of each foreign country, an indicator variable is defined that takes value 1 for RCAs greater than 1 and 0 in other cases. This means that with the new matrix of RCA indicators, the sum of the rows will show the diversity of the influence of each country on joint production.

$$RCA_{c'a'} = \left( P_{c'a'} \Big/ \sum_a P_{c'a} \right) \Big/ \left( \sum_c P_{ca'} \Big/ \sum_{c,a} P_{ca} \right)$$

where $c'$ indicates the country, $a'$ the area and $P$ the collaboration of this country in one area, being the sums referred to countries or areas.

In the same way as in Guevara et al. (2016), we built a Country *versus* Area matrix to measure the influence of countries on the proportion of DOIs in the cited references by area. This entailed using a fractional measure of the percentages of DOIs in relation to the number of participating countries. This indicates that if the percentage of DOIs in an article is zero, participating countries have no influence on the total. For example, in the case of an article in the area of Physics that has a DOI percentage of 0.4 with 4 different participating countries, the measure of influence of each country (for that article and the area of Physics) will be 0.4/4 = 0.1. Finally, we ranked the influence of each country using the proportion of DOIs per article, by adding all the columns (areas) of the matrix.

$$DOI_{ca} = \sum_i \frac{P_{cai}}{n_i}$$

where $i$ is the index of the paper with the proportion $P_{ca}$ and number of different countries of the authors $n_i$.

The creation of the rankings was based on the assumption that the ordinal positions of the countries in each area (resulting from either collaborative publications or the influence of the DOI) provides a score depending on the position.

The attributed score to the country $S_c$ ($c = 1,2,\ldots,n$) in the area $a$ ($a = 1,2,\ldots,23$) will depend on their position $j$ ($j = 1,2,\ldots,n$) and will be calculated according to the following formula:

$$S_c(a \lfloor j) = (n - j + 1)/n$$

where $n$ is the number of countries ($n = 196$). The country that occupies the first position has a score equal to 1, the score for the second position is $(n-1)/$ n and so on.

The total score for a country $c$ will be the sum of those obtained in each area ($a$). Therefore, if a country occupies the first position in all areas (maximum value), its total score will be 23.

Finally, we proceeded a graphical analysis in order to describe the dispersion of the most frequent collaborating countries and/or influential countries in the DOI attribution.

## Results

Before DOIs can be adopted in cited references, the article cited must, of course, have an identifier assigned, which is not common in the case of older articles. As mentioned above, several journals have made further efforts to carry out a DOI assignment. It should also be

**Table 3** Descriptive statistics of percentage of DOIs in the cited references of the articles (all, 2002 onwards and difference)—2012–2016

| Statistics | All references | 2002 onwards | Difference |
|---|---|---|---|
| Q1 | 35.3% | 41.2% | 5.9% |
| Median | 65.0% | 73.1% | 8.1% |
| Q3 | 80.8% | 87.7% | 6.9% |
| Mean | 56.1% | 61.6% | 5.6% |
| SD | 30.6% | 32.6% | 2.0% |
| Asym. coef | −0.65 | −0.80 | −0.15 |

noted that the use of this information in cited references is sometimes relatively unknown in the scholarly community, and may lead to differences, in view of the various factors that are analyzed.

Table 3 shows all the statistics with a higher percentage of DOIs in the articles, when the cited references are restricted to 2002 onwards. In addition, there is a difference of 0.15 between the asymmetry coefficients, which reveals a greater degree of skewness in the distribution of articles with restricted cited references. This negative skewness means that there is a stronger concentration, despite some increase of standard deviation, and that is also a sign of a tendency to strengthen the use of DOIs, particularly in the more recently cited references.

The same statistics displayed in Table 3 were employed to form the clusters of knowledge areas (as observed in Fig. 1). Hence, Table 4 provides the descriptive statistics for each of the clusters and in Fig. 2 the areas are sorted in ascending order of the cluster numbers, as well as the percentage of DOIs in the cited references of the articles. With regard to the percentage of DOIs in the cited references in Table 4, it should be noted that the greater the number of clusters, the greater the difference between comprehensive and restricted sets of cited references. This means it can be stated that the more established the attribution of DOIs is in the areas, and/or their use in cited references, the greater their effect with regard to the most recent references. For this reason, it is necessary to check in advance if the effect of the factors changes at the beginning or end of the period under analysis.

The first two clusters are comprised by the main areas of the soft sciences, with just one area for each one, and both being positively skewed. Cluster 1 (in Table 4), which is the area of Human Sciences, has the smallest number of articles (4695, and 2.1%), and presents at least 50% of the articles without DOIs in the cited references, with many outliers within the medium and high percentages (Fig. 2). Social Sciences is another exception, and forms Cluster 2, with at least 25% of the articles without DOIs in the cited references and 7.3% (16,536) of the articles. This is in line with previous studies that found low percentages for the availability of DOIs in articles from the social sciences, and even lower for humanities (Gorraiz et al. 2016; Gumpenberger et al. 2016). Although they found higher percentages, since they were dealing with source data, while we analyzed the cited references. This means that even though publishers are responsible for assigning DOIs to their articles, they might not require authors to use DOIs in their cited references during the editing process.

Cluster 3 comprises the other social science areas (Economics & Business), plus Agricultural Sciences and Plant & Animal Sciences, which are areas in which the local context is very important. Their distributions are more heterogeneous, and a wide dispersion is evident, with some negative skewness (Table 4 and Fig. 2). This cluster concentrates 22.2% (50,304) of the articles. Among all areas, Economics & Business is the only one that shows

**Table 4** Descriptive statistics of percentage of DOIs in the cited references of the articles (all and difference)—according to clusters of areas of knowledge—2012–2016

| Cluster | All references | | | | | | Difference % DOI (2002 onwards)—(all refs.) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 (%) | Median (%) | Q3 (%) | Mean (%) | SD (%) | Asym. coef | Q1 (%) | Median (%) | Q3 (%) | Mean (%) | SD (%) | Asym. coef |
| 1 | 0.0 | 0.0 | 5.6 | 4.7 | 10.0 | 3.45 | 0.0 | 0.0 | 1.1 | 1.1 | 2.6 | −0.35 |
| 2 | 0.0 | 11.8 | 44.8 | 23.8 | 26.9 | 0.84 | 0.0 | 2.5 | 5.2 | 2.3 | 2.3 | −0.05 |
| 3 | 16.3 | 42.8 | 64.6 | 40.9 | 27.8 | −0.03 | 4.3 | 5.9 | 6.3 | 4.7 | 2.2 | −0.13 |
| 4 | 39.9 | 60.7 | 74.1 | 55.2 | 25.2 | −0.63 | 5.5 | 7.2 | 7.6 | 5.5 | 2.6 | −0.12 |
| 5 | 65.7 | 78.4 | 86.8 | 72.9 | 20.6 | −1.57 | 7.8 | 7.4 | 6.1 | 6.2 | 0.7 | −0.34 |
| All areas | 45.7 | 62.0 | 74.7 | 58.2 | 22.8 | −0.76 | 5.9 | 6.6 | 6.3 | 5.4 | 1.6 | −0.23 |

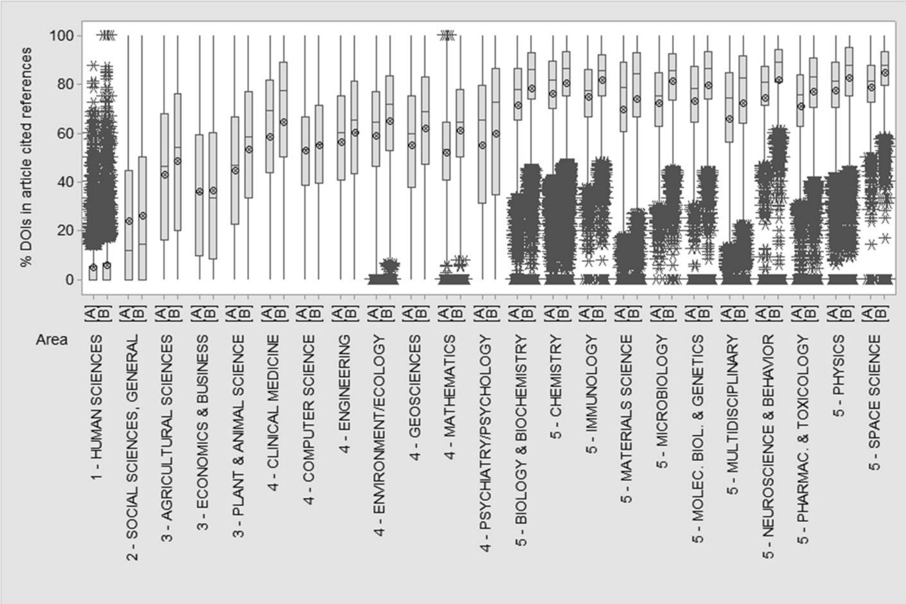


**Fig. 2** Distribution of percentage of DOIs in the cited references of the articles, with [A] referring to all and [B] to 2002 onwards—according to cluster and area of knowledge—2012–2016

a decline when the analysis is restricted to recent cited references. In contrast, Plant & Animal Sciences is the area that showed the highest increase in the median (11.7%).

In Cluster 4, there is some initial evidence of negative skewness and moderate dispersion (Table 4), and is formed of the following: Clinical Medicine, Computer Science, Engineering, Environment/Ecology, Geosciences, Mathematics and Psychiatry/Psychology (Fig. 2). All the areas in this group have a median higher than 55% of DOIs in both sets of cited references and Mathematics shows the highest increase in the third quartile (13.5%) when the comprehensive and restricted sets of cited references are compared. It aggregates 33.4% (75,607) of the articles, being Clinical Medicine the most prolific of all areas, with 17.9%.

Aggregating the eleven remaining areas Cluster 5 is the biggest in numbers of areas and articles (79,349 or 35.0%). These areas have the highest negative skewness (Table 4) and numerous outliers on the bottom and right-hand side of Fig. 2. The highest third quartile is shown by Chemistry (81.5%), when all the cited references are included, while Physics takes the lead (95.1%) when restricted to 2002 onwards. The highest mean is from Space Science—which for the comprehensive and restricted sets of cited references are 78.4% and 84.5% respectively.

## Influencing factors

The analysis of General Linear Models was conducted as a means of evaluating to what extent researchers are likely to attribute DOIs to cited references in journal articles, by comparing the factors' categories. It should be pointed out that the level of significance was always assumed to be 1%, which is worth mentioning as it might be assumed to be 5%.

The model uses one of the categories as a benchmark, for making a comparison with the others.

We decided to use Physics as a benchmark for the areas of knowledge. In the case of the comprehensive set (with all the cited references included), only Neuroscience & Behavior was found not to be significantly different from Physics, while all the other areas being less likely to use DOIs (Chemistry is the only exception with higher percentages, as can be observed in Fig. 2). Slight differences were observed with the restricted set (in the cited references from 2002 onwards): Chemistry and Molecular Biology & Genetics were not significantly different from Physics; in turn, Neuroscience & Behavior, as well as Chemistry, were more likely to use DOIs; while there was significantly less likelihood in all of the other areas of knowledge doing the same. Thus, over and above what was established in H1, not only are the soft sciences significantly different from the hard sciences, but there is significant heterogeneity even among hard sciences.

When attention is drawn to the other factors involved, Table 5 makes it possible to analyze the "type of collaboration", in which the category *single authorship* has the lowest frequency rate, as well as the median and mean percentages of DOIs in the cited references. Articles on *institutional* and *national collaboration* follow, with a medium level of median and mean percentages of DOIs in the cited references of the articles, although there is a greater increase of median in *institutional collaboration* when in the case of the restricted set of cited references. Finally, *international collaboration* has the highest percentages, as well as the second highest number of articles—just behind *national collaboration*. The probability test used *international collaboration* as a benchmark for the other categories, and only mentioned *national collaboration* as not significantly different from it, in the case of the comprehensive set. On the other hand, all the categories were found to be significantly different in the restricted set, which means that H1 is only confirmed in this case.

It could be argued that the median is clearly high when there is more than one author, and especially if there is a foreign author or newer cited references are being analyzed. This could be a sign that the adoption of DOIs in cited references is more established abroad. At the same time, it might be just a reflection of an earlier attribution of DOIs to articles, by publishers in some countries to the detriment of others (Boudry and Chartron 2017). As mentioned above, many domestic journals use the ABNT/NBR 6023 standard as citation style, but this only includes guidelines about DOIs for 2018. Additionally, students in Brazilian post-graduate scientific training programs are urged to conform to this standard in their dissertations. Another factor that should be noted, is that soft science areas in particular compel their students to comply with this standard as well. To sum up, these differences in the adoption of standards exist on a national, regional, institutional, departmental and individual level as well as between academic disciplines. Obviously, similar differences can also be found in other countries.

Another factor which deserves attention is the "status of the articles" in WoS (with regard to their openness), where *SciELO Gold* shows the lowest median percentages and *Green Accepted* the highest. In light of this, is worth taking note of some additional data not shown in Table 5. The articles in the SciELO journals are characterized by the following factors: (1) there is only 9.1% of *international collaboration* (with a median of DOIs of 27.8%), thus they are closely related to concerns in the domestic community; (2) some of the articles (mainly those from Clusters 2 and 3, which cover areas in a local context, have a low median percentage of DOIs (Clinical Medicine 0%, Agricultural Sciences 19.0%, Plant & Animal Sciences 20.0% and Social Sciences 0%). In contrast, the articles labeled *Green Accepted* not surprisingly have (1) a 98.0% rate of *international collaboration* (with median of DOIs of 82.8%); (2) 63.6% of the articles belong

**Table 5** Distribution of articles from 2012–2016, median and mean percentage of DOIs in the cited references (all and difference), and significance ($p$ value) of the factors' categories (all and from 2002 onwards)

| Factor | Category | Articles | Median | | Mean | | Significance test | |
|---|---|---|---|---|---|---|---|---|
| | | | All refs. (%) | Diff. (2002 onwards)—(all) (%) | All refs. (%) | Diff. (2002 onwards)—(all) (%) | All refs. | 2002 onwards |
| Type of collaboration | Single author | 8611 | 7.1 | 1.9 | 23.6 | 3.5 | 0.00 | 0.00 |
| | Institutional | 58,888 | 61.7 | 8.3 | 53.3 | 5.4 | 0.00 | 0.00 |
| | National | 88,760 | 60.0 | 7.7 | 51.6 | 5.4 | **0.146** | 0.00 |
| | International | 70,232 | 74.3 | 7.5 | 68.0 | 6.2 | – | – |
| Status of the article—closed access or open access route | SciELO Gold | 50,418 | 17.2 | 3.8 | 28.4 | 3.5 | 0.00 | 0.00 |
| | DOAJ Gold | 21,806 | 73.3 | 7.6 | 68.1 | 6.6 | 0.00 | 0.00 |
| | Other Gold | 383 | 75.0 | 6.3 | 68.3 | 5.3 | 0.00 | 0.00 |
| | Bronze | 14,868 | 74.1 | 8.5 | 70.0 | 7.7 | 0.00 | 0.00 |
| | Green Published | 6298 | 80.4 | 7.8 | 76.6 | 7.0 | 0.00 | 0.00 |
| | Green Accepted | 2122 | 82.7 | 7.3 | 80.5 | 6.7 | *0.027* | **0.112** |
| | Not open | 127,141 | 68.8 | 7.7 | 61.5 | 5.9 | – | – |
| Presence of DOI in the article? | No | 30,176 | 30.6 | 6.4 | 33.8 | 5.0 | 0.00 | **0.122** |
| | Yes | 196,315 | 68.8 | 7.9 | 59.5 | 5.7 | – | – |
| Mention of research funding? | No | 88,557 | 39.2 | 5.6 | 38.9 | 4.2 | 0.00 | 0.00 |
| | Yes | 137,934 | 73.3 | 7.7 | 67.1 | 6.5 | – | – |

Significance at the 1% level was determined for two decimal places, while at the 5% level the italics is applied, and bold was used to highlight those that were not significant

to Clinical Medicine (median of DOIs of 81.8%), Neuroscience & Behavior (84.3%), Biology & Biochemistry (84.3%) and Immunology (81.3%).

As can be seen from this contrasting situation, the area or presence of a foreign author can influence the extent to which DOIs are adopted in different ways, depending on the "open access route". This is why our analyses examine the interaction between factors' categories (as will be explained below). However, it is worth noting the apparent contradiction between the highest median percentages of the *Green Accepted* route, and the non- significant difference to its benchmark (*Not open*) in the restricted set. This is due to the fact that the standard error from this route is higher (data not shown), which means that despite having a higher median (or mean), the heterogeneity of the variable prevents it from differing significantly.

Table 5 also shows that the median and mean of the *Green Published* route follow the *Green Accepted* category very closely. However, it is worth noting that *Green Published* is more likely to assign DOIs in the cited references than the *Not open* category (in both reference sets), while in the case *Green Accepted,* we found there was only significance in the comprehensive set (at a rate of 2.7%). This means that those authors who make use of repositories to share their articles appear to be more careful about including DOIs in the references cited. The higher percentages of green routes corroborate with Jacsó's (2004) affirmation that preprint archives, "go out of their way to make as many cited references actionable as possible". From then until now, several types of repositories have been created, and increased their representativeness in the Brazilian national output. This could also explain why articles on Physics and Space Science had the highest level of DOIs in the cited references; it is because they belonged to the areas which created the first preprint server (arXiv) in 1991. However, currently this has spread to other areas, such as Mathematics and Engineering, which cover a significant proportion of WoS articles (Larivière et al. 2014).

Finally, Other Gold, DOAJ Gold and Bronze have median and mean percentages that are close. They all lie between the green routes and Not open, and show a significant likelihood of assigning DOIs to the cited references. Something that deserves attention is the fact that, although these categories are the equivalent of 17.9% of the total number of articles, their main contributors are in the areas of Cluster 5 (excluding Chemistry and Materials Science), where they represent 35.0%. Since these are the areas with the highest prevalence of DOIs in the cited references, it can be expected that they will influence these percentages.

The other factors studied were whether or not the article had a DOI assigned to it and if the acknowledgements make any mention of research funding. It can be seen that these factors appear in most of the articles, and that the median and mean percentage of DOIs is almost double when they are present. In other words, if the article has a DOI attributed to it, it will have more cited references with DOIs; moreover, the same can be said, if the article results from a funded research. To test their significance, we examined the "presence of DOI in the article" and the "mention of funding" (i. e. the presence of information about this) and used it as a benchmark. The result is that there is a lower likelihood that articles without DOIs will use DOIs in their cited references, although this was only found in the comprehensive set. Similarly, when the question of funding was not included, the likelihood is lower, but in this case to both reference sets.

As could be seen, all the factors had an influence on the percentage of DOIs in the cited references. However, with regard to our RQ1, it can be said that their influence is of a different degree of intensity, and depends on each of their categories.

### Interacting factors and temporal changes

When analyzing the other factors in each area of knowledge (Table 6), there is a very diverse scenario. As can be observed, no *p* values are attributed to Physics since this was taken as a benchmark. Just for the sake of space, the benchmarks for each factor are not given in the columns (*international collaboration* and *presence of DOI in the article* or *mention of research funding*).

It is clear that the areas that form contrasting clusters show a significance in all the factors/categories, compared with Physics, where there is a much greater difference from the resulting Clusters 1, 2 and 3. With regard to the other areas, the differences are significant in a more generalized way in the factor "presence of DOI in the article" (the only exception being Space Science in both sets) and the "mention of funding" (with the exception of Environment/Ecology in the comprehensive set). As was noted in the descriptive statistics (Table 5), the median and mean percentage of DOIs in the cited references of the articles was much higher when these factors were present.

The "type of collaboration" shown in Table 6 suggests that its influence is found to be more significant among Clusters 1, 2 and 3. In other words, the areas in these clusters have significantly different percentages of DOIs in their cited references, from all these types of collaboration, as *international collaboration* is taken as benchmark. This means that the same situation can also be confirmed just in a couple of areas of Clusters 4 (Environment/Ecology) and 5 (Multidisciplinary and Pharmacology & Toxicology). Thus, our H2 is confirmed since the areas of Clusters 1, 2 and 3, as well as Pharmacology & Toxicology, are those with lower percentage of international collaboration, and show that the presence of a foreign author is a decisive factor in the increase of the percentage of DOIs in the cited references.

The articles that form part of the "open access routes" show a different pattern among the areas (Table 7). In general, all the golden routes show significant differences with regard to non-open access articles, with the exception of: Economics & Business and Chemistry in the *Other Gold* route; Engineering and Multidisciplinary in the *SciELO Gold* route (it is important to note the low frequency of articles, as can be seen in Table 2). Clusters 1 and 3 show fewer areas with significant differences in the *Bronze* route, while the following are also not significant: Clinical Medicine in Cluster 3 and Biology & Biochemistry, Chemistry and Immunology in Cluster 5.

The *Green Published* route has some areas that do not differ significantly from Physics in Clusters 1, 4 and 5; in turn, to the *Green Accepted* route, the great majority of areas not differ significantly. The above-mentioned apparent contradiction between the higher median percentages mainly of the *Green Accepted* route and the insignificant difference to their benchmark (*Not open*) is detailed in the analysis of interaction with the areas (Table 7), where we observe few exceptions of areas that differ significantly from Physics.

All these analyses of interactions took note of the full set of articles, in the whole period of 2012–2016. The question of whether the interactions have changed in recent years was checked by replicating the GLM model at the beginning (2012–2013) and end (2015–2016) of the bienniums (i.e. periods of 2 years). We only examined the percentage of DOIs in the cited references of articles from 2002 onwards. Appendices 1 and 2 set out the significance (*p* value) of each biennium, as well as for the whole period (related to Tables 6 and 7, respectively). Table 8 is based on the results from Appendices 1 and 2, and summarizes the analysis of the areas. However, it ignores those areas

**Table 6** Significance (*p* value) of the factors (type of collaboration, presence of DOI in the article and mention of research funding) with regard to the areas of knowledge, considering the percentage of DOIs in the cited references of the articles (all and from 2002 onwards)

| Cluster | Area | Single author | | Inst. collab. | | Nat. collab. | | No DOI art. | | No funding | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | 2002 | All | 2002 | All | 2002 | All | 2002 | All | 2002 |
| 1 | HUMAN SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | SOCIAL SCIENCES, GENERAL | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | AGRICULTURAL SCIENCES | 0.00 | 0.00 | 0.00 | *0.018* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | ECONOMICS & BUSINESS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | PLANT & ANIMAL SCIENCE | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | CLINICAL MEDICINE | **0.286** | **0.128** | **0.052** | **0.746** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | COMPUTER SCIENCE | 0.00 | 0.00 | **0.084** | **0.257** | **0.140** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | ENGINEERING | **0.245** | 0.00 | **0.068** | *0.013* | **0.053** | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | ENVIRONMENT/ECOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.717** | *0.047* |
| | GEOSCIENCES | **0.256** | **0.074** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | MATHEMATICS | 0.00 | 0.00 | *0.015* | *0.048* | **0.117** | *0.026* | 0.00 | 0.00 | 0.00 | 0.00 |
| | PSYCHIATRY/PSYCHOLOGY | 0.00 | 0.00 | **0.730** | **0.501** | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | BIOLOGY & BIOCHEMISTRY | 0.00 | 0.00 | 0.00 | 0.01 | **0.080** | **0.720** | 0.00 | 0.00 | 0.00 | 0.00 |
| | CHEMISTRY | 0.00 | 0.00 | **0.179** | **0.345** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | IMMUNOLOGY | *0.031* | **0.163** | 0.00 | 0.00 | **0.416** | **0.198** | 0.00 | 0.00 | 0.00 | 0.00 |
| | MATERIALS SCIENCE | 0.01 | **0.077** | 0.00 | 0.00 | *0.030* | **0.508** | 0.00 | 0.00 | *0.012* | 0.00 |
| | MICROBIOLOGY | **0.824** | **0.095** | 0.00 | **0.271** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | MOLECULAR BIOLOGY & GENETICS | 0.00 | 0.00 | **0.205** | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | MULTIDISCIPLINARY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | NEUROSCIENCE & BEHAVIOR | **0.594** | **0.177** | 0.01 | 0.00 | 0.00 | **0.484** | 0.00 | 0.00 | 0.00 | 0.00 |
| | PHARMACOLOGY & TOXICOLOGY | 0.00 | 0.00 | *0.016* | *0.011* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | SPACE SCIENCE | **0.101** | *0.012* | *0.021* | **0.979** | 0.00 | **0.224** | **0.051** | **0.142** | 0.00 | 0.00 |
| | PHYSICS | – | – | – | – | – | – | – | – | – | – |

Significance at the 1% level was determined for two decimal places, while at the 5% level the italics is applied, and bold was used to highlight those that were not significant

**Table 7** Significance (*p* value) of the factor (open access route) with regard to the areas of knowledge, considering the percentage of DOIs in the cited references of the articles (all and from 2002 onwards)

| Cluster | Area | SciELO Gold | | DOAJ Gold | | Other Gold | | Bronze | | Green Publis. | | Green Accep. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | 2002 | All | 2002 | All | 2002 | All | 2002 | All | 2002 | All | 2002 |
| 1 | HUMAN SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.590** | **0.103** | **0.277** | **0.721** | – | – |
| 2 | SOCIAL SCIENCES, GENERAL | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | *0.040* | **0.065** |
| 3 | AGRICULTURAL SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | *0.038* | **0.692** | 0.00 | 0.00 | **0.278** | **0.061** |
| | ECONOMICS & BUSINESS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.099** | 0.00 | **0.057** | 0.00 | 0.00 | – | – |
| | PLANT & ANIMAL SCIENCE | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.076** | **0.495** | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | CLINICAL MEDICINE | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.220** | 0.00 | 0.00 | **0.129** | **0.463** |
| | COMPUTER SCIENCE | | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.080** | *0.045* |
| | ENGINEERING | 0.00 | **0.434** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.399** | **0.113** |
| | ENVIRONMENT/ECOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.310** | 0.00 | **0.833** | **0.766** |
| | GEOSCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.178** | *0.046* |
| | MATHEMATICS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | **0.079** | 0.00 | **0.083** | **0.406** |
| | PSYCHIATRY/PSYCHOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.536** | **0.118** | **0.264** |
| 5 | BIOLOGY & BIOCHEMISTRY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.523** | 0.00 | **0.597** | 0.00 | **0.774** | **0.569** |
| | CHEMISTRY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.299** | 0.00 | 0.00 | *0.017* | 0.00 | **0.636** | **0.845** |
| | IMMUNOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.394** | **0.923** | 0.00 | **0.594** | **0.490** | **0.622** |
| | MATERIALS SCIENCE | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.466** | **0.114** | **0.536** | **0.946** |
| | MICROBIOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | *0.026* | 0.01 | **0.482** | **0.307** |
| | MOLECULAR BIOLOGY & GENETICS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.397** | **0.490** | **0.630** | **0.382** |
| | MULTIDISCIPLINARY | **0.109** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | NEUROSCIENCE & BEHAVIOR | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | *0.016* | **0.061** | 0.00 | **0.097** | *0.049* | **0.409** |
| | PHARMACOLOGY & TOXICOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | *0.024* | 0.00 | 0.00 | **0.387** | **0.622** |
| | SPACE SCIENCE | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | – | – |
| | PHYSICS | – | – | – | – | – | – | – | – | – | – | – | – |

Significance at the 1% level was determined for two decimal places, while at the 5% level the italics is applied, and bold was used to highlight those that were not significant

**Table 8** Percentage of areas of knowledge in each cluster that change the significance ($p$ value) of its interactions with the categories of the factors, during the period, agreeing (agree) or altering (differ) the significance of the end of biennium compared with the whole period

| Factor | Category | Cluster | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | | 4 | | 5 | |
| | | Agree (%) | Differ (%) | Agree (%) | Differ (%) | Agree (%) | Differ (%) | Agree (%) | Differ (%) | Agree (%) | Differ (%) |
| Type of collaboration | Single author | | | | | 33 | | 14 | 29 | | 20 |
| | Institutional | 100 | | | | 33 | 33 | 29 | 57 | 20 | 40 |
| | National | 100 | | | 100 | 33 | | 43 | 14 | 40 | 30 |
| Presence of DOI in the article? | No | | 100 | | | | 67 | 14 | 29 | | 70 |
| Mention of research funding? | No | | | | | | 33 | | 43 | 10 | 40 |
| Open acces route | SciELO Gold | | | | | 67 | | 14 | 29 | 10 | |
| | DOAJ Gold | 100 | | | | | | 43 | | 40 | |
| | Other Gold | 100 | | | | | 33 | | | 20 | |
| | Bronze | 100 | | | 100 | | 100 | 14 | 29 | 20 | 30 |
| | Green Published | | | | | 67 | 33 | 43 | 14 | 10 | 30 |
| | Green Accepted | | | 100 | | 33 | 33 | 14 | 43 | 30 | |

that did not show changes in their significance between both bienniums and the whole period—which means that the influential factor has not changed over a period of time. In contrast, Table 8 shows the percentage of areas that change over time in each cluster, which are divided as follows: (a) when the period (biennium 2015–2016) ends, it has the same pattern of behavior ("Agree") as that observed for the whole period; or (b) in contrast, just at the beginning of the period (biennium 2012–2013), it has the same behavior but ends the period (biennium 2015–2016) by differing ("Differ") from the whole period.

Table 8 shows that in Cluster 1, Human Sciences maintained its behavior in both bienniums as observed throughout the period—with regard to s*ingle authored* articles and its benchmark (*international collaboration*). It is worth noting that *institutional* and *national collaboration* showed no significant difference regarding the percentage of DOIs in the references at the beginning of the period, but changed at the end, when it followed the pattern of the whole period. In marked contrast, in the case of articles where DOIs were absent, there was a significant difference at the beginning of the period, but this changed at the end and thus contradicted the significance observed about the whole period. This means that it makes no difference to the percentage of DOIs in the references, if the article includes a DOI or not, particularly at the end of the period and this suggests that the use of DOIs in the references for this area becomes independent of this factor. In view of these details about Cluster 1, Table 8 can be interpreted by stating that Human Science showed a general tendency to change its pattern of behavior in half of the factors' categories, but ended the period by following the pattern observed for the whole period.

The "presence of DOI in the article" showed to be independent from the use of DOIs in references. It can be observed in a general manner among the clusters, with Cluster 2 (Social Sciences) being the only exception. Other areas that are exceptions among the clusters are Economics & Business, Clinical Medicine, Geosciences, Mathematics, Psychiatry/Psychology, Materials Science, Microbiology and Multidisciplinary.

The factor "mention of research funding", in turn, has proved to be independent from the use of DOIs in references to one third of the areas of Cluster 3 (Plant & Animal Science), 43% of Cluster 4 (Computer Science, Engineering and Geosciences) and 40% of Cluster 5 (Molecular Biology & Genetics, Neuroscience & Behavior, Pharmacology & Toxicology and Space Science).

With regard to the "type of collaboration" factor, *institutional collaboration* was found to be not significantly different from *international collaboration* (its benchmark) at the end of the period. It can be said that the presence of a foreigner among the authors from the same institution, does not affect the probability that DOIs are used in references to the following areas: Agricultural Sciences in Cluster 3; Engineering and Mathematics in Cluster 4; Biology & Biochemistry, Molecular Biology & Genetics and Neuroscience & Behavior.

A similar behavior can be found in *national collaboration* with regard to *international collaboration* at the end of the period, where the presence of a foreigner within the authors from different Brazilian institutions does not influence the use of DOI in the references. This was observed in Social Sciences and Molecular Biology & Genetics. Finally, when we examined the s*ingle authored* articles, this change in the significance at the end of the period, with regard to *international collaboration*, is only observed in Biology & Biochemistry—but it should be remembered the very low frequency of *single authored* articles in this area, as shown in Table 2.

Among the “open access routes”, the pattern of behavior is quite different from the other factors previously analyzed, since in Table 8, it can be seen that there is greater agreement when the categories of the factors in the areas are compared. The *DOAJ Gold* route deserves attention, since many areas that do not show a significant difference at the beginning of the period, change their behavior at the end, and are in agreement with the whole period. This means that at the beginning of the period, the articles published in the DOAJ “fully open access” journals did not provide evidence that they used DOI in references with a greater probability, than with the *Not Open* access articles. This is the case with Cluster 1 (Human Sciences), 43% of areas from Cluster 4 (Clinical Medicine, Environment/Ecology and Mathematics) and 40% of Cluster 5 (Material Sciences, Molecular Biology & Genetics, Multidisciplinary and Space Science). The *Green Published* route shows the same pattern in 67% of the areas in Cluster 3 and 43% in Cluster 4.

In the *Bronze* route, most of the disagreements are concentrated at the end of the period, when compared with the whole period, in all the clusters, except for one. In Cluster 3, there is a curious situation: although the difference is not significant for the whole period, the opposite is the case in each of the bienniums. The same was observed with Immunology and Neuroscience & Behavior, from Cluster 5. In the *Green Accepted* route, Cluster 4 should be pointed out because of the disagreements at the end of the period, which apply to 43% of the areas (Computer Science, Geosciences and Psychiatry/Psychology), when compared with the whole period.

Finally, we checked the interaction of the other factors with the open access routes (Table 9). This revealed that *national collaboration* has no significance in all the golden routes in the restricted set (2002 onwards), while in the case of *institutional* and *single authorship* we only observe it to *Other Gold*. The analysis of the changes in behavior between open access routes and the other factors—according to "Appendix 3", to evaluate if interactions have changed in recent years—reveals fewer changes in *SciELO Gold*, *DOAJ Gold* and *Green Accepted* during the period, which shows that there was a significant difference in both bienniums resulting from the *national collaboration*, and conflicts with the observation made about the whole period..

In the case of *Bronze*, *single authorship* is the only type of collaboration that does not present significance with (Table 9), showing difference at the end of the period, with regard to the whole period ("Appendix 3"), since the “presence of DOI in the article” showed to be independent from the use of DOIs in the references.

In fact, all the open access routes show significance for the factor “presence of DOI in the article”, while in the case of “mention of research funding” all the routes in the restricted set are significant, with the exception of *Bronze* (Table 9). It is clear from "Appendix 3" that there are changes during the period, since the “presence of DOI in the article” becomes insignificant, while “mention of funding” is significant in both bienniums.

*Green Accepted* is able to interact with all types of collaboration in the restricted set, while the same is true for *Green Published*, but in the comprehensive set. Although there are many changes in both green routes in "Appendix 3", it should be remembered that these routes have very low frequencies, as is clear from Table 2, which is more fragile when the data are divided between two bienniums.

With regard to our RQ2, it can be stated that there is only a significantly different pattern between Clusters 1, 2 and 3 and other areas, when the "type of collaboration" factor is taken into account. A more generalized effect is exercised by the factors "presence of DOI

**Table 9** Significance (*p* value) of the open access routes with regard to other factors (type of collaboration, presence of DOI in the cited references of the articles and mention of research funding), considering the percentage of DOIs in the cited references of the articles (all and from 2002 onwards)

| Factor | Category | Single author | | Inst. collab. | | Nat. collab. | | No DOI art. | | No funding | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | 2002 | All | 2002 | All | 2002 | All | 2002 | All | 2002 |
| Status of the article—closed access or open access route | SciELO Gold | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.836** | 0.00 | 0.00 | 0.00 | 0.00 |
| | DOAJ Gold | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.455** | 0.00 | 0.00 | 0.00 | 0.00 |
| | Other Gold | **0.543** | 0.01 | 0.00 | **0.782** | 0.00 | **0.356** | 0.00 | 0.00 | 0.00 | 0.00 |
| | Bronze | *0.027* | **0.793** | 0.00 | *0.016* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.772** |
| | Green Published | 0.00 | **0.258** | *0.033* | **0.406** | 0.01 | *0.031* | 0.00 | 0.00 | 0.00 | 0.00 |
| | Green Accepted | 0.00 | *0.029* | **0.278** | 0.00 | **0.842** | *0.024* | 0.00 | 0.00 | 0.00 | 0.00 |
| | Not open | – | – | – | – | – | – | – | – | – | – |

Significance at the 1% level was determined for two decimal places, while at the 5% level the italics is applied, and bold was used to highlight those that were not significant

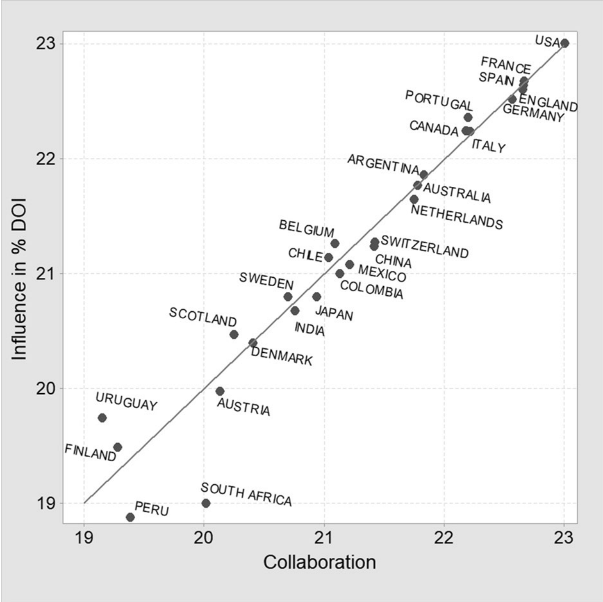


**Fig. 3** Scatter plot of the 27 principal countries collaborating with Brazilian institutions, based on rankings of collaborative publications and influence in the DOI attribution, considering all the areas of knowledge—2012–2016

in the article" and "mention of funding", as well as the golden categories of "open access routes", in almost all the areas. Bronze route highlighted Cluster 4 and many areas from Cluster 5, from the others, while the green routes were the only categories of factors that did not allow for any significant differentiation between the areas.

In a further point about RQ2, it can be stated that there is a similar pattern between SciELO and DOAJ Gold, when the "type of collaboration" factor is included, in contrast with the other routes, which cannot be differentiated from this factor. The same generalized effect of the factors "presence of DOI in the article" and "mention of funding" among the areas can be observed among the open access routes.

## The influence of countries

A set of the 27 countries that collaborate most and/or exert the greatest influence on the DOI attribution was selected for the following analysis, and this enabled us to answer our RQ3. As can be seen in Fig. 3, authors from institutions in the USA collaborate the most with authors from Brazilian institutions and these authors exert the main influence on the attribution of DOIs to cited references of the articles. A second group of countries is formed of the larger Western European countries (in terms of population size) and Canada, where there is also a trend for the attribution of DOIs to cited references—in at least 22 areas. Thus, there is a third group comprising those whose influence covers at least 21 areas, including four Latin American countries, Australia, China and some smaller countries from Western Europe. Finally, there are ten countries that have an influence on less than 21 areas, including five other smaller countries from Western Europe, Japan, India, South Africa and two more Latin American countries (Uruguay and Peru).

The linear dispersion shows that these variables are clearly correlated, while the diagonal reference line highlights the position of collaborative publications with countries like Uruguay, where the influence on DOI attribution is extended to more areas than is the case with its collaborations. This can be observed to a lesser extent in collaborative publications with Finland, Scotland, Sweden, Chile, Belgium and Portugal. In contrast, Peru and South Africa show that their influence is below their level of collaboration, which can be said to a lesser extent to Austria, Japan, Colombia, Mexico, China, Switzerland and the Netherlands.

## Conclusion

More than dealing with metadata, our study treats bibliographic records as abstractions of the documents themselves, which in turn can be regarded as artifacts of communication (Borgman and Furner 2002) with features (as expressed by metadata) that are embedded in their contexts of use. Alternatively, as Brown and Duguid (1995) argue; "we need to see the way documents have served not simply to write, but also to underwrite social interactions; not simply to communicate, but also to coordinate social practices". In this way, the assessment of the presence of DOIs in cited references serves as an indicator of openness and can enhance the open availability of citation data by making it possible to link them with documents for different uses.

The use of DOIs in cited references is a crucial factor in finding a match between the citation and the cited article, which can be of great value for citation indexes. Because of greater openness of citation-related data and the advent of Altmetrics, several analyses have been conducted that are based on data retrieved through DOIs, which is a feature of growing importance for specialists in scientometric studies. These showed that several bibliometric studies have investigated DOIs on the basis of different datasets (retrieved from other databases, at different time periods, etc.), but considering the articles, which meant that it was not possible to compare them with our findings. Although we believe that the presence of DOIs in the cited references is influenced by the percentage of attribution of DOIs to articles by the publishers, our literature review revealed differences among areas and countries.

The factors of influence revealed that there is a more pronounced tendency to use DOIs in the cited references to the hard sciences, but we even observed a clear heterogeneity among them (which confirms our first hypothesis). The factor "type of collaboration" illustrated the fact that the presence of an author from an institution outside Brazil is a key determinant in the use of DOIs, as well as when comparing multiple with single authorship. Nevertheless, the interaction of this factor with the areas of knowledge made clear that international collaboration differs significantly from the other types, for example in the soft sciences and other areas such as Agricultural Sciences, Environment & Ecology, Multidisciplinary and Plant & Animal Sciences. Hence our second hypothesis was also confirmed. In addition, we found a significantly higher percentage of DOIs in cited references of "articles with DOI assignment", as well as with "mention of research funding", even in the case of an interaction with the areas of knowledge.

It was noted as well that articles published in the green routes had the highest percentages of DOIs in cited references—in relation to articles *Not open*—among the "open access routes", followed by the golden routes (*DOAJ* and *Other*). Despite this, when analyzed in each area, the golden routes (*DOAJ* and *Other*) and Bronze take the lead and the green routes seem not to differ to most of the areas. The articles published in journals that are also indexed in SciELO had lower percentages, which suggests that the practice still needs to be consolidated among the publishers and authors in a national context.

The temporal analysis of the interacting factors showed that the *international collaboration* ceased to be an influential factor at the end of the period in the case some specific areas of knowledge. This is applied to *institutional collaboration*, particularly in the hard sciences (Agricultural Sciences, Biology & Biochemistry, Engineering, Mathematics, Molecular Biology & Genetics and Neuroscience & Behavior) and *national collaboration* (Social Sciences and Molecular Biology & Genetics).

"Mention of research funding" also ceased to be a factor of influence for the use of DOIs in references at the end of the period for the following: Computer Science, Engineering, Geosciences, Molecular Biology & Genetics, Neuroscience & Behavior, Pharmacology & Toxicology, Plant & Animal Science and Space Science.

With regard to publications in *DOAJ Gold*, there was no difference in the use of DOIs in the references to *Not open* access articles, but the situation was reversed at the end of the period (for Clinical Medicine, Environment/Ecology, Human Sciences, Material Sciences, Mathematics, Molecular Biology & Genetics, Multidisciplinary and Space Science).

Finally, the analysis of co-publications with authors from the most frequently collaborating countries showed that, practically, the presence of all of them have a similar influence on the DOI attribution in cited references. We discovered that the following three countries showed more (Uruguay) or less (Peru and South Africa) influence on the attribution of DOIs, which signals in what a way the analysis should be undertaken in future.

From the perspective of making advances in providing greater openness to science, by increasing the number of different forms of publication, with different versions of the same documents and formatting languages, it is essential to encourage the pervasive use of DOIs. In light of this, studies on the availability and usability of DOIs can assist in the greater use of DOI in the flow of scientific information in the digital era and ensure greater consistency when conducting bibliometric analysis.

**Acknowledgements** This work would not have been possible without the financial support of São Paulo Research Foundation (FAPESP) Grant #2012/00255-6 and CNPq (Research Productivity Grant #311237/2019-3).

## Appendix 1

See Table 10.

**Table 10** Significance ($p$ value) of the factors (type of collaboration, presence of DOI in the article and mention of research funding) with regard to the areas of knowledge, considering the percentage of DOIs in the cited references of the articles (2002 onwards): for the whole period (12–16), beginning (12–13) and end (15–16) bienniums

| Cluster | Área | Single Author | | | Institutional collaboration | | | National collaboration | | | No DOI in the article | | | No funding | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| 1 | HUMAN SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | **0.263** | 0.00 | 0.00 | **0.686** | 0.00 | 0.00 | 0.00 | **0.195** | 0.00 | – | 0.00 |
| 2 | SOCIAL SCIENCES, GENERAL | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.01 | **0.203** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | AGRICULTURAL SCIENCES | 0.00 | 0.00 | 0.00 | *0.018* | 0.00 | **0.753** | 0.00 | **0.305** | 0.00 | 0.00 | 0.00 | **0.670** | 0.00 | 0.00 | 0.01 |
| | ECONOMICS & BUSINESS | 0.00 | **0.889** | 0.00 | 0.00 | **0.948** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | PLANT & ANIMAL SCIENCE | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.272** | 0.00 | 0.00 | **0.517** |
| 4 | CLINICAL MEDICINE | **0.128** | **0.243** | 0.00 | **0.746** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| | COMPUTER SCIENCE | 0.00 | 0.00 | 0.00 | **0.257** | 0.00 | **0.189** | 0.00 | **0.516** | 0.00 | 0.00 | **0.062** | **0.322** | 0.00 | 0.00 | **0.628** |
| | ENGINEERING | 0.00 | **0.204** | 0.04 | 0.01 | 0.00 | **0.282** | 0.01 | **0.147** | 0.02 | 0.00 | 0.00 | **0.515** | 0.00 | 0.00 | **0.616** |
| | ENVIRONMENT/ECOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | **0.124** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.075** | 0.03 | 0.05 | 0.00 | 0.00 |
| | GEOSCIENCES | **0.074** | **0.616** | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.077** |
| | MATHEMATICS | 0.00 | 0.04 | 0.03 | 0.05 | **0.134** | **0.149** | 0.03 | **0.129** | **0.413** | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 |
| | PSYCHIATRY/PSYCHOLOGY | 0.00 | 0.00 | 0.00 | **0.501** | 0.00 | 0.00 | 0.01 | **0.380** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

**Table 10** (Continued)

| Cluster | Área | Single Author | | | Institutional collaboration | | | National collaboration | | | No DOI in the article | | | No funding | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| 5 | BIOLOGY & BIOCHEMISTRY | 0.00 | 0.00 | **0.294** | 0.01 | 0.00 | **0.492** | **0.720** | **0.103** | **0.354** | 0.00 | **0.767** | **0.731** | 0.00 | 0.00 | 0.00 |
| | CHEMISTRY | 0.00 | 0.00 | 0.04 | **0.345** | 0.00 | 0.00 | 0.00 | **0.687** | 0.00 | 0.00 | 0.00 | **0.967** | 0.00 | 0.00 | 0.00 |
| | IMMUNOLOGY | **0.163** | **0.664** | **0.334** | 0.00 | 0.00 | 0.00 | **0.198** | **0.222** | 0.00 | 0.00 | 0.00 | **0.696** | 0.00 | 0.00 | 0.03 |
| | MATERIALS SCIENCE | **0.077** | **0.093** | 0.00 | 0.00 | **0.353** | 0.00 | **0.508** | **0.439** | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | **0.114** | 0.00 |
| | MICROBIOLOGY | **0.095** | **0.256** | **0.687** | **0.271** | 0.00 | **0.419** | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | MOLECULAR BIOLOGY & GENETICS | 0.00 | 0.01 | 0.05 | 0.01 | **0.879** | **0.696** | 0.00 | 0.00 | **0.115** | 0.00 | 0.00 | **0.134** | 0.00 | 0.00 | **0.967** |
| | MULTIDISCIPLINARY | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.232** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | NEUROSCIENCE & BEHAVIOR | **0.177** | **0.415** | **0.941** | 0.00 | 0.00 | **0.100** | **0.484** | 0.04 | **0.076** | 0.00 | 0.00 | **0.145** | 0.00 | 0.00 | **0.089** |
| | PHARMACOLOGY & TOXICOLOGY | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.485** | 0.00 | 0.00 | **0.247** |
| | SPACE SCIENCE | 0.01 | 0.02 | 0.01 | **0.979** | **0.225** | **0.353** | **0.224** | 0.03 | **0.329** | **0.142** | **0.597** | 0.00 | 0.00 | 0.00 | **0.816** |
| | PHYSICS | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |

Significance at the 5% level was determined for two decimal places, while bold and three decimal places were used to highlight those that were not significant. The figures only included the percentage of DOIs in the cited references of the articles from 2002 onwards

## Appendix 2

See Table 11.

**Table 11** Significance (*p* value) of the factor (open access route) with regard to the areas of knowledge, considering the percentage of DOIs in the cited references of the articles (2002 onwards): for the whole period (12–16), beginning (12–13) and end (15–16) bienniums

| Cluster | Área | SciELO Gold | | | DOAJ Gold | | | Other Gold | | | Bronze | | | Green Published | | | Green Accepted | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| 1 | HUMAN SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | **0.262** | 0.00 | 0.00 | **0.735** | 0.00 | **0.103** | – | **0.890** | **0.721** | **1.000** | **0.259** | – | – | – |
| 2 | SOCIAL SCIENCES, GENERAL | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.749** | 0.00 | 0.00 | 0.02 | **0.065** | 0.01 | **0.855** |
| 3 | AGRICULTURAL SCIENCES | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.692** | 0.00 | 0.00 | 0.00 | **0.917** | 0.00 | **0.061** | 0.00 | **0.988** |
| | ECONOMICS & BUSINESS | 0.00 | **0.666** | 0.00 | 0.00 | 0.00 | 0.01 | **0.099** | **0.318** | 0.03 | **0.057** | 0.02 | 0.00 | 0.00 | **0.187** | 0.00 | – | – | – |
| | PLANT & ANIMAL SCIENCE | 0.00 | **0.330** | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | **0.495** | 0.00 | 0.00 | 0.00 | 0.00 | **0.425** | 0.00 | 0.00 | **0.160** |

**Table 11** (continued)

| Cluster | Área | SciELO Gold | | | DOAJ Gold | | | Other Gold | | | Bronze | | | Green Published | | | Green Accepted | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| 4 | CLINICAL MEDICINE | 0.00 | 0.00 | 0.00 | 0.00 | **0.986** | 0.00 | 0.00 | 0.00 | 0.00 | **0.220** | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | **0.463** | **0.237** | **0.183** |
| | COMPUTER SCIENCE | – | – | – | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.04 | **0.966** |
| | ENGINEERING | **0.434** | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | **0.365** | 0.00 | **0.113** | **0.075** | **0.172** |
| | ENVIRON-MENT/ECOL-OGY | 0.00 | 0.00 | 0.00 | 0.00 | **0.495** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.383** | 0.00 | **0.766** | 0.04 | **0.372** |
| | GEOSCIENCES | 0.00 | **0.059** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.03 | **0.896** |
| | MATHEMAT-ICS | 0.00 | 0.00 | 0.00 | 0.00 | **0.669** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.921** | 0.00 | 0.00 | **0.119** | 0.01 | **0.406** | **0.188** | **0.736** |
| | PSYCHIATRY/PSYCHOL-OGY | 0.00 | 0.00 | **0.950** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.174** | **0.536** | **0.757** | 0.04 | **0.264** | **0.094** | 0.03 |
| 5 | BIOLOGY & BIOCHEMIS-TRY | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.057** | 0.00 | 0.01 | **0.459** | **0.569** | **0.071** | **0.757** |
| | CHEMISTRY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.299** | 0.00 | **0.066** | 0.00 | **0.149** | 0.00 | 0.00 | **0.171** | **0.525** | **0.845** | **0.095** | **0.339** |
| | IMMUNOLOGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.923** | 0.00 | 0.03 | **0.594** | **0.868** | **0.486** | **0.622** | **0.118** | **0.328** |
| | MATERIALS SCIENCE | 0.00 | 0.00 | 0.00 | 0.00 | **0.105** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.179** | 0.00 | **0.114** | **0.838** | **0.089** | **0.946** | **0.615** | **0.995** |
| | MICROBIOL-OGY | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.05 | **0.307** | 0.02 | **0.890** |
| | MOLECULAR BIOLOGY & GENETICS | 0.00 | 0.00 | 0.00 | 0.00 | **0.085** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.490** | **0.840** | **0.657** | **0.382** | 0.04 | **0.765** |

**Table 11** (continued)

| Cluster | Área | SciELO Gold | | | DOAJ Gold | | | Other Gold | | | Bronze | | | Green Published | | | Green Accepted | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| | MULTIDISCIPLINARY | 0.00 | 0.02 | 0.00 | 0.00 | **0.964** | 0.00 | 0.00 | **0.300** | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | **0.160** | 0.00 | 0.00 | 0.00 | 0.00 |
| | NEUROSCIENCE & BEHAVIOR | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.061** | 0.00 | 0.00 | **0.097** | **0.318** | 0.01 | **0.409** | **0.168** | **0.078** |
| | PHARMACOLOGY & TOXICOLOGY | 0.00 | **0.850** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | **0.622** | 0.04 | **0.072** |
| | SPACE SCIENCE | 0.00 | 0.00 | 0.00 | 0.00 | **0.903** | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.03 | 0.00 | – | – | – |
| | PHYSICS | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |

Significance at the 5% level was determined for two decimal places, while bold and three decimal places were used to highlight those that were not significant. The figures only included the percentage of DOIs in the cited references of the articles from 2002 onwards

## Appendix 3

See Table 12.

**Table 12** Significance (*p* value) of the factors (type of collaboration, presence of DOI in the cited references of the articles and mention of research funding) with regard to the open access routes, considering the percentage of DOIs in the cited references of the articles (2002 onwards): for the whole period (12–16), beginning (12–13) and end (15–16) bienniums

| Factor | Category | Single Author | | | Inst. collab. | | | Nat. collab. | | | No DOI art. | | | No funding | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 | 12–16 | 12–13 | 15–16 |
| Status of the article—closed access or open access route | SciELO Gold | 0.00 | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | **0.836** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | DOAJ Gold | 0.00 | **0.858** | 0.00 | 0.00 | 0.00 | 0.00 | **0.455** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Other Gold | 0.01 | **0.170** | **0.192** | **0.782** | **0.293** | **0.471** | **0.356** | **0.321** | **0.980** | 0.00 | **0.063** | 0.00 | 0.00 | **0.105** | 0.00 |
| | Bronze | **0.793** | 0.03 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | **0.053** | 0.01 | 0.00 | 0.00 | **0.587** | **0.772** | 0.00 | 0.00 |
| | Green Published | **0.258** | 0.01 | **0.081** | **0.406** | **0.092** | **0.739** | 0.03 | 0.02 | 0.00 | 0.00 | 0.00 | **0.067** | 0.00 | **0.195** | 0.00 |
| | Green Accepted | 0.03 | **0.053** | | 0.00 | 0.00 | **0.378** | 0.02 | **0.425** | **0.065** | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | **0.05** |
| | Not open | – | – | – | – | – | – | – | – | – | – | – | – | – | – | |

Significance at the 5% level was determined for two decimal places, while bold and three decimal places were used to highlight those that were not significant. The figures only included the percentage of DOIs in the cited references of the articles from 2002 onwards

## References

Albagli, S., Clínio, A., & Raychtock, S. (2014). Ciência Aberta: correntes interpretativas e tipos de ação. *Liinc em Revista, 14*, 15. https://doi.org/10.18617/liinc.v10i2.749.

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*(1), 2–72. https://doi.org/10.1002/aris.1440360102.

Boudry, C., & Chartron, G. (2017). Availability of digital object identifiers in publications archived by PubMed. *Scientometrics, 110*, 1453–1469. https://doi.org/10.1007/s11192-016-2225-6.

Brown, J. S., & Duguid, P. (1995). The social life of documents. *First Monday*, *1*(1). Retrieved August 10, 2020, from https://firstmonday.org/ojs/index.php/fm/article/download/466/387.

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., & Wittenburg, P. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. https://doi.org/10.2777/1524.

Costas, R. (2017). General discussion on the most relevant characteristics of research infrastructures for scientometrics. In R. Mugnaini, A. Fujino, & N. Y. Kobashi (Eds.), *Bibliometrics and scientometrics in Brazil: Scientific research assessment infrastructure in the era of Big Data* (pp. 43–65). ECA/USP: São Paulo.

Crossref. (2020). *Crossref annual report & fact file 2018–19*. https://doi.org/10.13003/y8ygwm5.

Digiampietri, L., Mugnaini, R., Trucolo, C., Delgado, K., Mena-Chalco, J., & Köhler, A. (2019). Geographic and disciplinary distribution of the Brazilian's PHD community. *Brazilian Journal of Information Science: Research Trends, 13*(4), 113–131. https://doi.org/10.36311/1981-1640.2019.v13n4.07.

Fasae, J. K., & Oriogu, C. D. (2018). Digital object identifier and their use in accessing online scholarly materials in Africa. *Library Philosophy and Practice*, *1*. Retrieved August 10, 2020, from https://digitalcommons.unl.edu/libphilprac/1785.

Gibbs, W. W. (1995). Lost science in the third world. *Scientific American, 273*(2), 92–99. https://doi.org/10.1038/scientificamerican0895-92.

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*(1), 98–109. https://doi.org/10.1016/j.joi.2015.11.008.

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics, 109*(3), 1695–1709. https://doi.org/10.1007/s11192-016-2125-9.

Gumpenberger, C., Sorz, J., Wieland, M., & Gorraiz, J. (2016). Humanities and social sciences in the bibliometric spotlight—Research output analysis at the University of Vienna and considerations for increasing visibility. *Research Evaluation, 25*(3), 271–278. https://doi.org/10.1093/reseval/rvw013.

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics, 118*, 177–214. https://doi.org/10.1007/s11192-018-2958-5.

Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics, 108*(1), 413–423. https://doi.org/10.1007/s11192-016-1910-9.

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE, 10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495.

Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics, 113*(3), 1551–1571. https://doi.org/10.1007/s11192-017-2535-3.

International DOI Foundation. (2019). Factsheet. Key facts on digital object identifier system. Retrieved August 10, 2020, from https://www.doi.org/factsheets/DOIKeyFacts.html.

Jacsó, P. (2004). Link-enabled cited references. *Online Information Review, 28*(4), 306–311. https://doi.org/10.1108/14684520410553804.

Juty, N., Wimalaratne, S. M., Soiland-Reyes, S., Kunze, J., Goble, C. A., & Clark, T. (2020). Unique, persistent, resolvable: Identifiers as the foundation of FAIR. *Data Intelligence*. https://doi.org/10.1162/dint_a_00025.

Kikkawa, J., Takaku, M., & Yoshikane, F. (2016). DOI links on Wikipedia. In A. Morishima, A. Rauber, & C. Liew (Eds.), *Digital libraries: Knowledge, information, and data in an open access society. ICADL 2016. Lecture notes in computer science* (Vol. 10075). Cham: Springer. https://doi.org/10.1007/978-3-319-49304-6_40.

Kraft, A., Dreyer, B., Löwe, P., & Ziedorn, F. (2017). 14 Years of PID services at the German National Library of Science and Technology (TIB): Connected frameworks, research data and lessons learned from a national research library perspective. *Data Science Journal, 16*, 36. https://doi.org/10.5334/dsj-2017-036.

Larivière, V., Sugimoto, C. R., Macaluso, B., Milojevic, S., Cronin, B., & Thelwall, M. (2014). arXiv E-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology, 65*(6), 1157–1169. https://doi.org/10.1002/asi.23044.

Martín, S. G. (2013). El DOI en las revistas científicas del portal SciELO. *Palabra Clave (La Plata), 3*(1), 12–29.

Meneghini, R., & Packer, A. L. (2007). Is there science beyond English? *EMBO Reports, 8*(2), 112–116.

Montanari, F., & Packer, A. L. (2014). Criteria for the selection of journals to index and publish in the SciELO Network collections. In A. L. Packer, N. Cop, A. Luccisano, A. Ramalho, & E. Spinak (Eds.), *SciELO—15 years of open access: An analytic study of open access and scholarly communication* (pp. 67–80). Paris: UNESCO. https://doi.org/10.7476/9789230012373.

Mugnaini, R., Damaceno, R. J. P., Digiampietri, L. A., & Mena-Chalco, J. P. (2019). Panorama da produção científica do Brasil além da indexação: uma análise exploratória da comunicação em periódicos. *Transinformação, 31*, e190033. https://doi.org/10.1590/2318-0889201931e190033.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., et al. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ, 6*, e4375.

Rodríguez-Yunta, L. (2013). Pensando en digital: uso del DOI en revistas españolas, una asignatura aún pendiente. *Anuario ThinkEPI, 7*, 164–168.

Rubim, I. C., & Braganholo, V. (2017). Detecting referential inconsistencies in electronic CV datasets. *Journal of Brazilian Computer Society, 23*(1), 3. https://doi.org/10.1186/s13173-017-0052-0.

Ruiz, M. A. (2011). Countries emerging as major scientific powers. *Revista brasileira de hematologia e hemoterapia, 33*(3), 169–171. https://doi.org/10.5581/1516-8484.20110043.

Sugimoto, C. R., Murray, D. S., & Larivière, V. (2018). Open citations to open science. *ISSI Blog*. Retrieved August 10, 2020, from http://issi-society.org/blog/posts/2018/april/open-citations-to-open-science/.

van Eck, N. J., Waltman, L., Larivière, V., & Sugimoto, C. R. (2018). Crossref as a new source of citation data: A comparison with Web of Science and Scopus. *CWTS Blog*. Retrieved August 10, 2020, from https://www.cwts.nl/blog?article=n-r2s234.

Wang, W., Deng, L., You, B., Zhang, P., & Chen, Y. (2017). Digital object identifier and its use in core Chinese academic journals: A Chinese perspective. *Learned Publishing, 31*, 149–154. https://doi.org/10.1002/leap.1137.

Wang, X., Cui, Y., Xu, S., & Hu, Z. (2018). The state and evolution of Gold open access: A country and discipline level analysis. *ASLIB Journal of Information Management, 70*(5), 573–584. https://doi.org/10.1108/AJIM-02-2018-0023.

Wouters, P., Ràfols, I., Oancea, A., Kamerlin, S. C. L., Holbrook, J. B., & Jacob, M. (2019). Indicator frameworks for fostering open knowledge practices in science and scholarship. *Report of the Expert Group on Indicators for Researchers' Engagement with Open Science*. https://doi.org/10.2777/445286.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics, 101*(2), 1491–1513. https://doi.org/10.1007/s11192-014-1264-0.

## Affiliations

**Rogério Mugnaini[1] · Grischa Fraumann[1,2] · Esteban F. Tuesta[1] · Abel L. Packer[3]**

Grischa Fraumann
grischa.fraumann@tib.eu

Esteban F. Tuesta
tuesta@usp.br

Abel L. Packer
abel.packer@scielo.org

1 University of São Paulo, São Paulo, Brazil

2 TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

3 SciELO/FAPESP Program, São Paulo, Brazil