



Availability of digital object identifiers in publications archived by PubMed

Christophe Boudry, Ghislaine Chartron

► To cite this version:

Christophe Boudry, Ghislaine Chartron. Availability of digital object identifiers in publications archived by PubMed. *Scientometrics*, Springer Verlag, 2017, 10.1007/s11192-016-2225-6. hal-01444105

HAL Id: hal-01444105

<https://hal.archives-ouvertes.fr/hal-01444105>

Submitted on 23 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title page

Author

Christophe Boudry, PhD^{1,2,3}, Ghislaine Chartron, PhD³

Article Title

Availability of digital object identifiers in publications archived by PubMed

Affiliations

¹Normandie Univ, UNICAEN, Média Normandie. Caen, France

²URFIST, Ecole Nationale des Chartes, PSL Research University. Paris, France

³Laboratoire “Dispositifs d’Information et de Communication à l’Ère Numérique”,

EA7339, Conservatoire National des Arts et Métiers, Paris, France

Corresponding Author

Christophe Boudry

Média Normandie, Université de Caen Normandie

Esplanade de la Paix

CS 14032, 14032 Caen Cedex 5, France

christophe.boudry@enc-sorbonne.fr

Abstract

Digital Object Identifiers (DOIs) were launched in 1997 to facilitate the long-term access and identification of objects in digital environments. The objective of the present investigation is to assess the DOI availability of articles in biomedical journals indexed in the PubMed database and to complete this investigation with a geographical analysis of journals by the country of publisher. Articles were randomly selected from PubMed using their PubMed Identifier (PMID) and were downloaded from and processed through developed Hypertext Preprocessor language (PHP) scripts. The first part of the analysis focuses on the period 1966-2015 (50 years). Of the 496665 articles studied over this period, 201055 have DOIs (40.48%). Results showed that the percentage of articles with DOIs began to increase for articles published in the 2000s, with spectacular growth in the years 2002-2003, then reached a peak in 2015. Data on countries showed that some countries gradually implemented DOIs over the period 1966 to 2015 (the United States, the United Kingdom, and the Netherlands), while some did not (Russia, the Czech Republic, and Romania). The second part of the analysis focuses on the year 2015 and includes 268790 articles published in 2015, randomly selected to evaluate the current implementation of DOIs. In 2015, 86.42% of articles had DOIs. The geographical analysis of countries of publishers showed that some countries (Russia, Thailand, and Ukraine) still assigned few DOIs to articles in 2015. Thus, if the scientific community aims to increase the number and the usefulness of services rendered by DOIs, efforts must be made to generalize their use by all persons involved in scientific publication, particularly publishers.

Keywords

Digital object identifier – PubMed - Country – Publisher – Biology – Medicine – Publication – Journal article

Introduction

In the last few decades, access to scientific information has almost entirely migrated to a digital environment. This changing environment requires persistent and reliable unique identifiers to manage access to digital resources. The most common identifier on the web is the Uniform Resource Locator (URL), which provides the location of resources. URLs have many limitations.

They work properly only if the resources are not removed, and do not provide a means of associating meta information (e.g. file format) about the resource which could help to identify intellectual property rights (Carnevale and Aronsky 2007; Ducut et al. 2008; Park et al. 2011; Wagner et al. 2009). In biology and medicine, the PubMed Identifier (PMID) has long been operating to identify and access articles via the PubMed database. However, the scope of PMID is limited because it only includes journals indexed in PubMed in the field of biology and medicine and is functional only in this database. The lack of similar systems in other scientific fields, and therefore the lack of normalization, resulted in the establishment of the Digital Object Identifier (DOI) in 1997 (Paskin 1999). This system, managed by the International DOI Foundation (IDF) (Chandrakar 2006; “Digital Object Identifier System Handbook” 2016; Paskin 1999, 2010; Simmonds 1999) is now an ISO standard (“ISO 26324:2012 - Information and documentation - Digital object identifier system” 2012) and has been identified as one of the best choices among those complying with the definition of a digital identifier from a comparative evaluation based on seven criteria (Khedmatgozar and Alipour-Hafezi 2015).

A DOI name consists of two parts (Sidman and Davidson 2001; Simmonds 1999). The first part consists of a prefix corresponding to a unique numeric string beginning with the numeral 10 assigned by the International DOI Foundation or by DOI registration agencies (e.g. CrossRef for scientific publishers) to the registrant that submitted the information about the digital object (publishers in the case of scientific publications). The second part consists of a suffix corresponding to an alphanumeric string or series of strings assigned by the publisher and is used internally to identify the digital object.

The system operates in such a way that when a DOI is first assigned to an object, the publisher registers its identity and current network location in the DOI system. Incoming requests for the DOI are resolved to the appropriate URL. When the object moves, the publisher of that contents simply updates its DOI record with the new URL. Hence, accessing digital objects using DOIs ensures seamless and correct routing to the new and correct location of objects (Sidman and Davidson 2001) and enables permanent and unambiguous identification and access to digital objects. One of the main benefits of referencing an object by a DOI versus a simple URL is persistence. When an object moves, its URL changes but its DOI remains the same. Furthermore, using DOIs instead of URLs allows multiple resolution. In addition to the DOI and URL,

deposited metadata may describe multiple formats of an article (pdf, html, XML) and several locations (e.g. publisher's website, PubMed Central). Therefore, end users can choose their preferred format or download location (DeRisi et al. 2003). The DOI can also be applied at any level of granularity (to a whole book or to chapters, illustrations or tables) or any file type (text, image, audio-video) (Wang 2007).

With the establishment of DataCite in 2009, whose aims are to facilitate access to research data, to increase the acceptance of data publication, and to support data archiving (Brase et al. 2015; Honor et al. 2016; Neumann and Brase 2014), DOIs can also be assigned to research data. During the five past years, DataCite has grown into a global consortium and has assigned over four million DOIs to scientific datasets (Brase et al. 2015).

DOIs offer other advantages for all those involved in the scientific community:

For publishers, paying one annual fee allows access to articles via DOI links and functions even if content moves or changes ownership. DOIs also protect the copyright of published material (Rosenblatt 1997). The adoption of DOIs by publishers increases the discoverability of resources by search engines, increasing the traffic on websites: the number of site views for documents with a DOI were higher by an average of 66% using CrossRef compared to documents without the DOI service (Sieck 2003). Consequently, publishers increase their revenue. For example, for a medium-sized book publisher, the return on investment using the DOI could be as much as 12 times the cost of DOI implementation (Sieck 2003).

For libraries, DOI links can extend access to content not owned by the library and increase usage of acquired electronic resources for end-users at no charge (Wang 2007).

For researchers, DOIs provide access to over 19 million scientific objects available from over 20,000 journals and from over 1,500 publishers and companies ("crossref.org" 2016) by simply using a copy/paste of the DOI in a DOI resolver (e.g. <https://dx.doi.org/>) or using bibliographic tools such as Zotero ("Zotero | Home" 2016). Use of DOIs enriches end-user activity and the scholarly research process, and also guarantees persistent access and citability of articles.

For research funders, now that most of them require researchers to make their articles available in open access with different specifications (licenses and embargo periods), they need to track

output from hundreds or even thousands of publishers. Thus, DOIscan offer a reliable way to help these institutions with their compliance processes (“Crossref initiatives will support reporting to funders | Research Information” 2016).

DOIs can also be used in scientific production evaluation. For example, the National Council for Scientific and Technological Development (CNPq) in Brazil signed an agreement with ISI Thomson Reuters to allow access and online viewing of the number of citations in the Web of Science database for the articles registered in the Curriculum Lattes (a virtual platform integrating curricula databases, research groups, and institutions) with their respective DOIs (Braile 2011).

Altmetrics are social web metrics for academic publications incorporating a number of variables such as view count, downloads, and comments in order to measure the impact of articles (Galligan and Dyas-Correia 2013; González-Valiente et al. 2016; Rasmussen and Andersen 2013; Sud and Thelwall 2013). The DOI is an essential element for most sites or tools using altmetrics (Galligan and Dyas-Correia 2013). For example, altmetric.com (“Altmetric” 2016), in order to disambiguate mentions of articles, searching for identifiers such as DOIs in web pages, PLOS Article-Level Metrics (“ALM” 2016) or PlumX (“Plumx” 2015) can create reports from one DOI or from a list of DOIs. Furthermore, following the success of a DOI event tracker (“CrossRef’s DOI Event Tracker Pilot – Crossref Blog” 2015; Tolwinski 2015), CrossRef will launch, in early 2017, the service CrossRef Event Data (“Crossref Event Data” 2016), which is an open data service based on DOIs. It will track activity surrounding a research work from potentially any web source where an event is associated with a DOI. This service will help researchers to follow the activity surrounding their articles or research data. Publishers will be able to use the service to track the dissemination of published articles to discover where they are being discussed, bookmarked, and linked. This will also help publishers to answer questions such as, “Is use growing over time?” or, “Which articles or subjects areas are being seen more than others?”, considering multiple web sites rather than only their own web site (Tolwinski 2015). DOIs are also used as tools in some studies addressing altmetrics, e.g. ensuring the link between bibliometric databases and altmetric sites (Haustein et al. 2015) or identifying articles (Haustein et al. 2014; Thelwall et al. 2013).

As mentioned before, DOIs and the development of their related services are increasingly important for academia at different levels (articles, research data, etc.). But the same is true for bibliometrics and altmetrics, which justifies studies focusing on the assignment of DOIs to scientific articles.

To the best of our knowledge, only one study has evaluated the implementation of DOIs in the scientific literature, using Scopus - which indexes several databases, including Embase, Medline or Biobase (Valderrama-Zurián et al. 2015) - and Web of Science Core Collection over the period 2005-2014 (Gorraiz et al. 2016). This study showed that “there are still journals with a large number of items still lacking DOIs in 2014” and this “should be alarming for the corresponding editors and should give them reason to enhance the formal quality and visibility of their journals...”. The authors also encourage scientists to review their publication strategies and to favour publication channels with established DOI assignments to result in a higher web presence and visibility”.

The objectives of the present study are:

- to assess the DOI availability of articles in biomedical journals indexed in the PubMed database, year by year, over the period 1966-2015 (50-year retrospective study),
- to evaluate the percentage of articles with DOIs in the PubMed database, published in 2015,
- to complete these investigations with a geographical analysis of the countries and continents of publishers in order to identify the countries and continents which assign the most and the least DOIs to articles. To the best of our knowledge, this type of analysis has never been performed.
- to identify the most productive journals in the field of biology and medicine which did not yet assign DOIs in 2015.

Materials and methods

The search for papers to be included in this study was carried out from March 22nd to April 20th, 2016 using the PubMed database (<http://www.ncbi.nlm.nih.gov/pubmed>), developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM). PubMed was chosen because it is the most widely used bibliographic database in

medicine (Falagas et al. 2008), and the country of publishers can be easily identified using the specific field “Country of Publication”. However, contrary to other bibliographic databases (e.g. Scopus or Web of Science), any search tool is effective for distinguishing references with DOIs from others in PubMed. In PubMed, the interrogation of the field “Article Identifier” (which includes article identifiers submitted by journal publishers such as DOI) with the wildcard operator (e.g. 10* to extract all articles with a DOI) uses only the first 600 variations, because PubMed searches for the first 600 variations of a truncated term. It was then impossible to extract all references with DOIs using a query to assess the availability of DOIs in PubMed articles. Therefore, references were randomly selected from PubMed according to their PubMed identifier (PMID) using a Hypertext Preprocessor language (PHP) function that generates random integers (the PHP random number generation function called `mt_rand (min,max)`, which returns an integer between min and max). The minimum value used was 1, and the maximum value 26995803, corresponding to the higher PMID assigned in PubMed at the time of the experiment was done. Then, the randomly generated integer was used to query PubMed by PMID using Efetch Entrez Programming Utilities, in order to extract the reference of the corresponding article. Data were downloaded from PubMed in Extensible Markup Language (XML) and were processed through developed PHP scripts. They were then imported to Microsoft Excel 2013 (Microsoft, Redmond, USA) for data processing as done previously (Boudry et al. 2016; Boudry and Mouriaux 2015). The analysis was limited to the publication type “Journal Article”. For each article, the existence of a DOI was verified, then the date of publication, the name of the journal, and the country of publishers were analyzed. When the country of publisher was absent in the reference, it was determined using the NLM catalog (“Home - NLM Catalog - NCBI” 2016). England, Scotland, Northern Ireland, and Wales were grouped into the United Kingdom. Countries were clustered by their continent according to the United Nations classification (“United Nations Statistics Division- Standard Country and Area Codes Classifications (M49)” 2016).

In order to assess implementation of DOIs year by year over the period 1966-2015, an analysis of 496665 randomly selected articles published from 1966 to 2015 was done, corresponding to 2.29% of the articles included in PubMed (the total number of articles in PubMed was 21680488 on March 22nd 2016 for the same period). This first analysis was called "retrospective analysis" in this paper.

A second analysis of 268790 randomly selected articles published in 2015 was done to evaluate the current implementation of DOIs (representing 21.68% of articles published in 2015 in the PubMed database). This second analysis was called “current analysis” in this paper. The year 2016 was not included because it was incomplete when the analysis was done, and a bias could have been introduced in the results when including journals with undefined frequencies of publication.

Results

Implementation of DOIs (retrospective analysis 1966-2015)

Among the 496665 articles studied, the number of articles with DOIs was 201055, corresponding to 40.78 % of articles with a DOI. Fig. 1 shows the percentage of articles with DOIs for articles published over the period 1966 to 2015. The percentage of articles with DOIs began to increase for articles published around the year 2000, with spectacular growth in the years 2002-2003, then reached a peak in 2015.

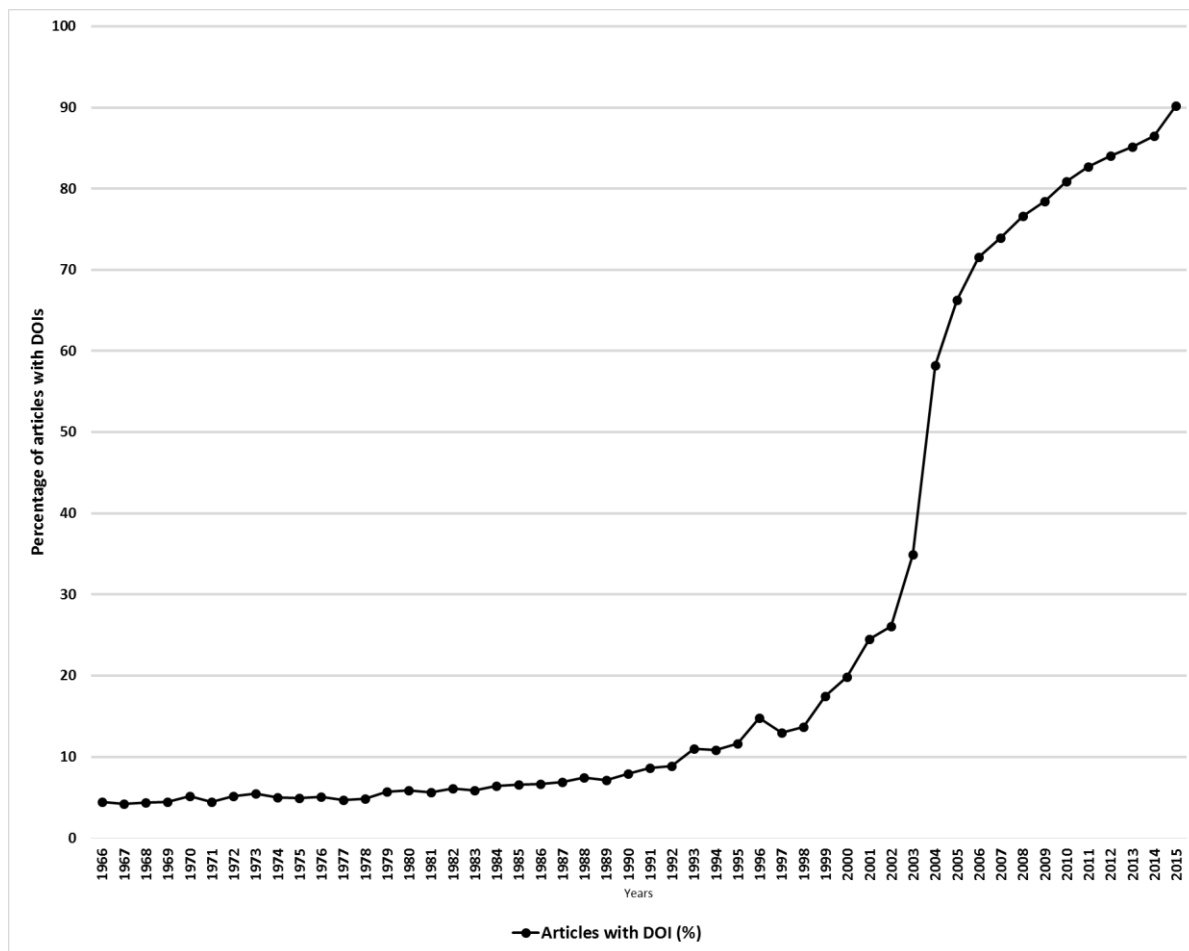


Fig. 1: Percentage of articles with DOI in the PubMed database for articles published over the period 1966-2015 (retrospective analysis)

Implementation of DOIs: geographical analysis of countries and continents of publishers (retrospective analysis 1966-2015)

In the period 1966-2015, we identified 115 countries of publishers. As shown in Table 1, among the 50 most productive countries of publishers, publishers from South Korea assigned the highest percentage of DOIs to articles, and publishers from seven countries assigned less than 1% of DOIs to articles they published over the period 1966-2015.

Table 1 Number of articles published, number of articles with DOIs and percentage of articles with DOIs for the 50 most productive countries of publishers over the period 1966-2015 (retrospective analysis)

| Country of publisher | Number of articles | Number of articles with DOIs | Percentage of articles with DOIs | Country of publisher (continued) | Number of articles | Number of articles with DOIs | Percentage of articles with DOIs |
|-------------------------------|--------------------|------------------------------|----------------------------------|----------------------------------|--------------------|------------------------------|----------------------------------|
| United States | 212668 | 98240 | 46.19 | Norway | 1658 | 297 | 17.91 |
| United Kingdom | 90291 | 51217 | 56.72 | Korea (South) | 1534 | 1180 | 76.92 |
| Netherlands | 27079 | 12770 | 47.16 | Greece | 1499 | 303 | 20.21 |
| Germany (1990-) | 19595 | 13117 | 66.94 | Austria | 1411 | 414 | 29.34 |
| Japan | 15713 | 1718 | 10.93 | Hungary | 1173 | 176 | 15.00 |
| Switzerland | 11588 | 4787 | 41.31 | Romania | 1001 | 10 | 1.00 |
| France | 11463 | 2068 | 18.04 | Mexico | 810 | 7 | 0.86 |
| West Germany (-1990) | 10717 | 917 | 8.56 | South Africa | 796 | 44 | 5.53 |
| USSR (-1991) | 9931 | 0 | 0.00 | Yugoslavia (-2006) | 681 | 0 | 0.00 |
| China | 8254 | 1041 | 12.61 | Egypt | 664 | 438 | 65.96 |
| Italy | 8083 | 807 | 9.98 | Czech Republic (1993-) | 659 | 16 | 2.43 |
| Denmark | 5760 | 1438 | 24.97 | Thailand | 643 | 0 | 0.00 |
| Canada | 5135 | 792 | 15.42 | Israel | 620 | 6 | 0.97 |
| Poland | 4716 | 270 | 5.73 | Iran | 601 | 161 | 26.79 |
| India | 4547 | 1768 | 38.88 | Finland | 594 | 9 | 1.52 |
| Ireland | 4294 | 2354 | 54.82 | Turkey | 588 | 126 | 21.43 |
| Australia | 4241 | 1408 | 33.20 | Bulgaria | 586 | 74 | 12.63 |
| Spain | 3571 | 542 | 15.18 | Singapore | 478 | 112 | 23.43 |
| Russia (1991-) | 3467 | 13 | 0.37 | Pakistan | 450 | 99 | 22.00 |
| Sweden | 2965 | 389 | 13.12 | Ukraine (1991-) | 443 | 0 | 0.00 |
| Brazil | 2712 | 602 | 22.20 | Argentina | 431 | 30 | 6.96 |
| East Germany (-1990) | 2595 | 36 | 1.39 | Chile | 347 | 88 | 25.36 |
| Czechoslovakia (-1992) | 2043 | 0 | 0.00 | Saudi Arabia | 307 | 47 | 15.31 |
| Belgium | 1822 | 97 | 5.32 | United Arab Emirates | 301 | 41 | 13.62 |
| New Zealand | 1760 | 710 | 40.34 | Croatia (1991-) | 276 | 20 | 7.25 |

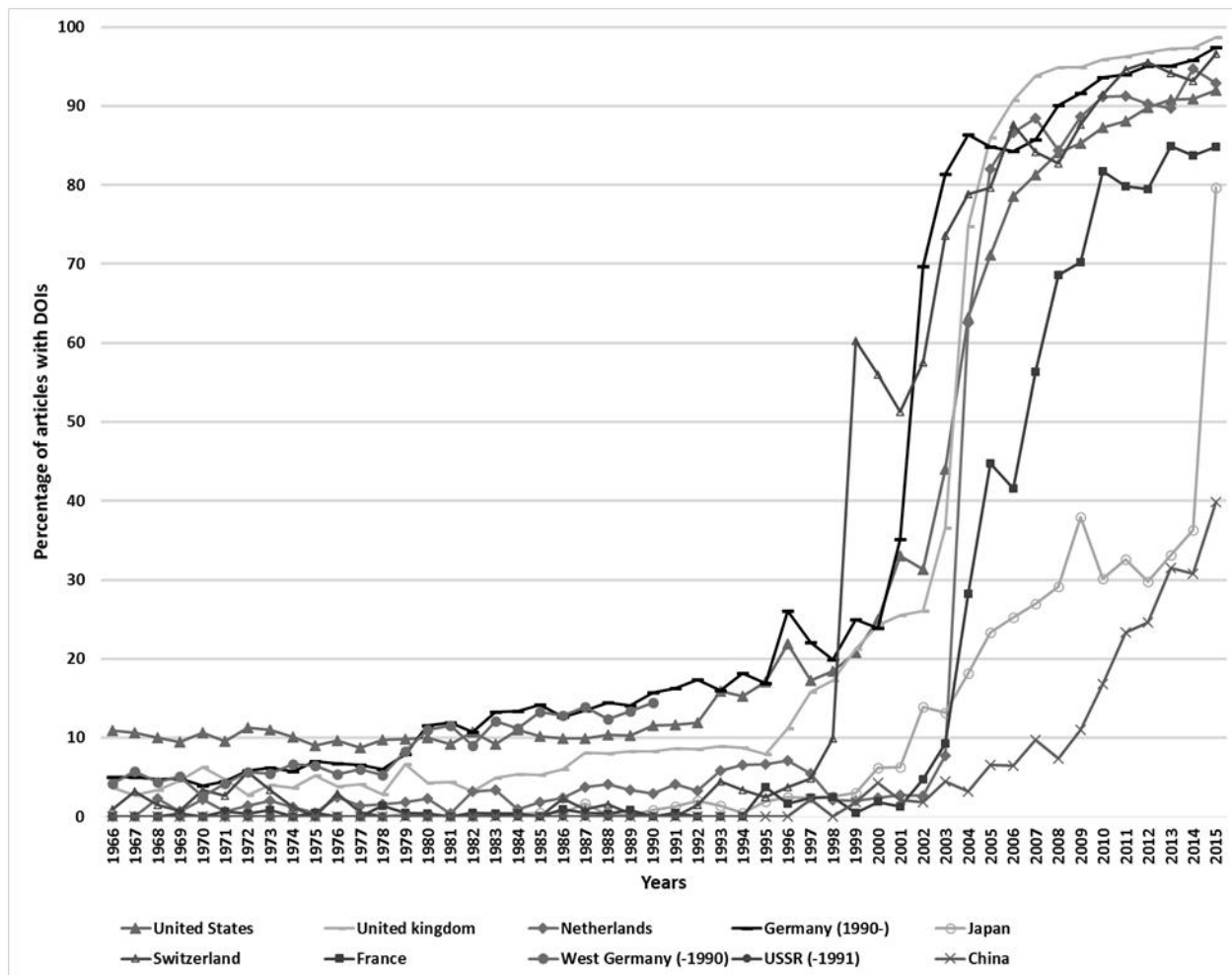


Fig. 2: Percentage of articles with DOIs in the PubMed database published over the period 1966-2015: The ten most productive countries of publishers (retrospective analysis)

The analysis of the percentage of articles with DOIs published by the 10 most productive countries (Fig 2) identified three groups of countries. The first includes 6 countries (France, Germany (1990-), Switzerland, the United Kingdom, the Netherlands, and the United States) where publishers massively assigned DOIs in the 2000s, to reach 85% to 99% of articles with DOIs in 2015. The second, with two countries (Japan, and China), began to assign DOIs after the first group and lagged behind the first group in 2015 (from 40% to 80% of articles with DOIs). The last group, with two countries, West Germany (-1990) and the USSR (-1991), had logically stopped their production in 1990 and 1991, when they ceased to exist, respectively. It is important to note that only publishers from three countries (the United States, Germany, and to a

lesser extent the United Kingdom) made an effort to assign DOIs retrospectively before 1997, the year DOIs were launched by the IDF.

Table 2 Number of articles published, number of articles with DOIs, and percentage of articles with DOIs for continents of publishers over the period 1966-2015 (retrospective analysis)

| Continent of publisher | Number of articles | Number of articles with DOIs | Percentage of articles with DOIs |
|--|--------------------|------------------------------|----------------------------------|
| Europe | 230740 | 92217 | 39.97 |
| North America | 217803 | 99032 | 45.47 |
| Asia | 34756 | 6355 | 18.28 |
| Oceania | 6029 | 2118 | 35.13 |
| Latin America and the Caribbean | 4917 | 749 | 15.23 |
| Africa | 2420 | 584 | 24.13 |
| World | 496665 | 201055 | 40.48 |

For continents, over the period 1966-2015 the percentage of articles with DOIs varied from 15.23% (Latin America and the Caribbean) to 45.47% (North America) of articles with DOIs (Table 2).

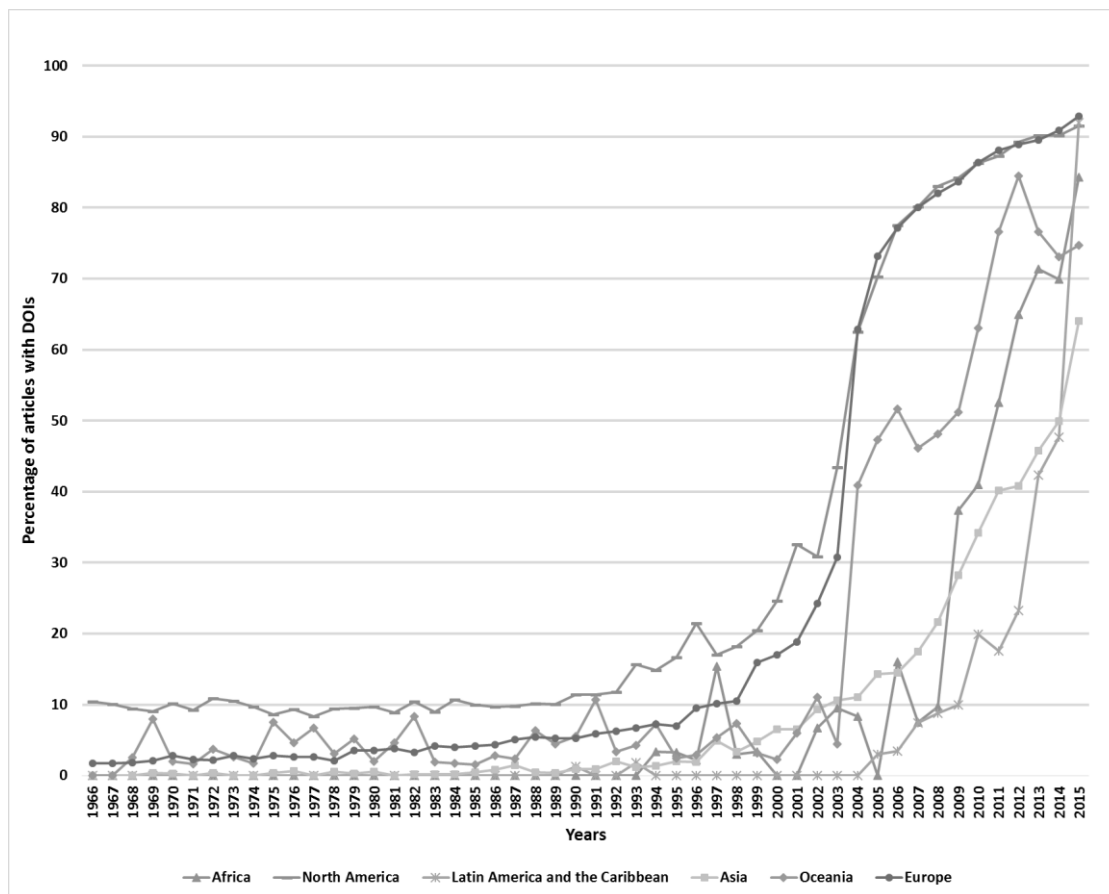


Fig. 3: Percentage of articles with DOIs in the PubMed database published from 1966 to 2015: continents of publishers

As shown in Fig. 3, three continents massively assigned DOIs in the 2000s: North America, Europe, and to a lesser extent Oceania. Asia, Latin America, the Caribbean, and Africa began later, around 2007/2008. Publishers from only two continents (North America and Europe) made an effort to assign DOIs retrospectively before 1997.

Implementation of DOIs in 2015: geographical analysis of countries and continents of publisher (current analysis)

To evaluate the current implementation of DOIs, 268,790 articles published in 2015 were analyzed (corresponding to 21.68% of the total number of articles published in 2015 referenced in PubMed). Of these articles, 232,281 have DOIs, corresponding to 86.42% of the articles. We identified 79 countries of publishers. The 268,790 articles were published in 7,600 journals. Out of

these 7600 journals, 6445 (84.80%) assigned DOIs, 1155 (15.22%) did not assign DOIs to their articles.

Results of countries and continents of publishers are presented in Table 3 and 4. Korea (South), Norway, and Uganda (but with only 2 journals for the two latter countries) were the countries where the percentage of journals using DOIs were the highest. Seven of the 50 most productive countries of publishers assigned less than 10% of DOIs to articles in 2015 (Israel, Malaysia, Mexico, Russia, Thailand, Ukraine and Serbia). In Russia, only 2 journals of 59 assigned DOIs to articles, representing 2.09% of articles with DOIs. Asia, in 2015, was the continent where the publishers implemented by far the least number of DOIs to articles (58.52%), while Europe led with 90.32% of articles with DOIs.

Table 3 Number of articles published, number of articles with DOIs, percentage of articles with DOIs, number of journals, number of journals with DOIs, and percentage of journals with DOIs in the 50 most productive countries of publishers for the year 2015 (current analysis)

| Country of publisher | Number of articles | Number of articles with DOIs | Percentage of articles with DOIs | Number of journals | Number of journals with DOIs | Percentage of journals with DOIs |
|----------------------|--------------------|------------------------------|----------------------------------|--------------------|------------------------------|----------------------------------|
| United States | 118291 | 105529 | 89.21 | 2810 | 2427 | 86.37 |
| United Kingdom | 62944 | 61364 | 97.49 | 1724 | 1667 | 96.69 |
| Netherlands | 14242 | 12893 | 90.53 | 479 | 439 | 91.65 |
| Germany | 12014 | 11559 | 96.21 | 393 | 373 | 94.91 |
| Switzerland | 10722 | 10266 | 95.75 | 248 | 236 | 95.16 |
| China | 6169 | 1934 | 31.35 | 140 | 73 | 52.14 |
| India | 4838 | 4124 | 85.24 | 168 | 145 | 86.31 |
| Japan | 4497 | 3403 | 75.67 | 164 | 122 | 74.39 |
| New Zealand | 3490 | 2495 | 71.49 | 124 | 119 | 95.97 |
| Italy | 2840 | 1417 | 49.89 | 113 | 63 | 55.75 |
| France | 2594 | 2186 | 84.27 | 96 | 84 | 87.50 |
| Canada | 2065 | 1127 | 54.58 | 87 | 40 | 45.98 |
| Australia | 2043 | 1509 | 73.86 | 83 | 66 | 79.52 |
| Korea (South) | 1826 | 1733 | 94.91 | 120 | 119 | 99.17 |
| Ireland | 1762 | 1710 | 97.05 | 38 | 36 | 94.74 |
| Brazil | 1660 | 1646 | 99.16 | 60 | 59 | 98.33 |
| Russia | 1582 | 33 | 2.09 | 57 | 2 | 3.51 |
| Poland | 1412 | 875 | 61.97 | 72 | 47 | 65.28 |
| Spain | 1329 | 1011 | 76.07 | 63 | 49 | 77.78 |
| Iran | 1245 | 622 | 49.96 | 78 | 28 | 35.90 |
| Greece | 1044 | 395 | 37.84 | 19 | 9 | 47.37 |
| Sweden | 892 | 237 | 26.57 | 29 | 9 | 31.03 |
| Denmark | 854 | 596 | 69.79 | 30 | 26 | 86.67 |
| Turkey | 813 | 387 | 47.60 | 31 | 19 | 61.29 |
| Thailand | 654 | 1 | 0.15 | 4 | 1 | 25.00 |
| Egypt | 540 | 510 | 94.44 | 68 | 65 | 95.59 |
| Uganda | 492 | 492 | 100.00 | 2 | 2 | 100.00 |
| Pakistan | 466 | 74 | 15.88 | 10 | 4 | 40.00 |
| United Arab Emirates | 445 | 117 | 26.29 | 38 | 5 | 13.16 |
| Czech Republic | 364 | 48 | 13.19 | 20 | 5 | 25.00 |
| Romania | 333 | 45 | 13.51 | 13 | 3 | 23.08 |
| Bulgaria | 312 | 284 | 91.03 | 8 | 6 | 75.00 |
| Austria | 302 | 299 | 99.01 | 19 | 17 | 89.47 |
| Hungary | 274 | 220 | 80.29 | 16 | 9 | 56.25 |

| | | | | | | |
|-------------------------------|-----|-----|--------|----|----|--------|
| Singapore | 266 | 197 | 74.06 | 14 | 12 | 85.71 |
| Belgium | 263 | 60 | 22.81 | 14 | 2 | 14.29 |
| Saudi Arabia | 242 | 185 | 76.45 | 9 | 7 | 77.78 |
| Mexico | 232 | 21 | 9.05 | 9 | 3 | 33.33 |
| Croatia | 204 | 31 | 15.20 | 12 | 4 | 33.33 |
| Ukraine | 193 | 0 | 0.00 | 8 | 0 | 0.00 |
| Israel | 161 | 9 | 5.59 | 6 | 1 | 16.67 |
| South Africa | 153 | 30 | 19.61 | 10 | 5 | 50.00 |
| Slovakia | 142 | 47 | 33.10 | 5 | 2 | 40.00 |
| Argentina | 127 | 34 | 26.77 | 8 | 4 | 50.00 |
| Norway | 123 | 123 | 100.00 | 2 | 2 | 100.00 |
| Malaysia | 108 | 0 | 0.00 | 7 | 0 | 0.00 |
| Serbia | 107 | 0 | 0.00 | 3 | 0 | 0.00 |
| Bosnia and Herzegovina | 105 | 95 | 90.48 | 6 | 5 | 83.33 |
| Chile | 97 | 65 | 67.01 | 4 | 3 | 75.00 |
| Finland | 93 | 15 | 16.13 | 2 | 1 | 50.00 |

Table 4 Number of articles published, number of articles with DOIs, percentage of articles with DOIs, number of journals, number of journals with DOIs, and percentage of journals with DOIs by continent of publishers for the year 2015 (current analysis)

| Continent of publisher | Number of articles | Number of articles with DOIs | Percentage of articles with DOIs | Number of journals | Number of journals with DOIs | Percentage of journals with DOIs |
|--|--------------------|------------------------------|----------------------------------|--------------------|------------------------------|----------------------------------|
| North America | 120356 | 106656 | 88.62 | 2897 | 2467 | 85.16 |
| Europe | 117199 | 105856 | 90.32 | 3500 | 3100 | 88.57 |
| Asia | 22061 | 12910 | 58.52 | 806 | 545 | 67.62 |
| Oceania | 5533 | 4004 | 72.37 | 207 | 185 | 89.37 |
| Latin America and the Caribbean | 2256 | 1783 | 79.03 | 94 | 73 | 77.66 |
| Africa | 1385 | 1072 | 77.40 | 96 | 75 | 78.13 |
| World | 268790 | 232281 | 86.42 | 7600 | 6445 | 84.80 |

Table 5 lists the 20 most productive journals for articles published in 2015 (current analysis), based on the number of articles published, that did not assign DOIs to their articles.

Table 5 The 20 most productive journals lacking DOIs: name of the journal, number of articles published, name of the publisher, country of publisher, ISSN and language of publication for articles published in 2015 (current analysis)

| Journal | Number of articles | Publisher | Country of publisher | ISSN | Language of publication |
|--|--------------------|--|----------------------|-----------|-------------------------|
| Int J Clin Exp Med | 1187 | e-Century Publishing | United States | 1940-5901 | English |
| Int J Clin Exp Pathol | 786 | e-Century Publishing | United States | 1936-2625 | English |
| J Nanosci Nanotechnol | 546 | American Scientific Publisher | United States | 1533-4880 | English |
| Asian Pac. J. Cancer Prev. | 504 | Asian Pacific Organization for Cancer Prevention | Thailand | 1513-7368 | English |
| Stud Health Technol Inform | 466 | IOS Press | Netherlands | 0926-9630 | English |
| Anticancer Res. | 379 | International Institute of Anticancer Research | Greece | 1791-7530 | English |
| Nippon Rinsho | 282 | Nippon Rinsho Co | Japan | 0047-1852 | Japanese |
| Eur Rev Med Pharmacol Sci | 257 | Verduci Editore | Italy | 2284-0729 | English |
| Zhonghua Yi Xue Za Zhi | 224 | Zhonghua yi xue hui | China | 0376-2491 | Chinese |
| Zhongguo Zhong Yao Za Zhi | 220 | Zhongguo yao xue hui | China | 1001-5302 | Chinese |
| Lakartidningen | 204 | Sveriges Lakarforbund | Sweden | 1652-7518 | Swedish |
| Rev Med Suisse | 204 | Médecine et Hygiène | Switzerland | 1660-9379 | French |
| Guang Pu Xue Yu Guang Pu Fen Xi | 195 | Beijing da xue chu ban she | China | 1000-0593 | Chinese |
| Ugeskr. Laeg. | 180 | Den Alm Danske Laegerforening | Denmark | 1603-6824 | Danish |
| Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi | 170 | Lin chuang er bi yan hou tou jing wai ke za zhi bain ji bu | China | 1001-1781 | Chinese |
| Am Surg | 167 | Southeastern Surgical Congress | United States | 1555-9823 | English |
| Huan Jing Ke Xue | 166 | Ke xue zhu ban she | China | 0250-3301 | Chinese |
| Ned Tijdschr Geneesk | 162 | Bohn Stafleu van Loghum | Netherlands | 1876-8784 | Dutch |
| Nat Prod Commun | 161 | Natural Product Communications | United States | 1934-578X | English |
| J Pak Med Assoc | 155 | Pakistan Medical Association | Pakistan | 0030-9982 | English |

| | | | | | |
|---|-----|--|---------------------|-----------|------------------|
| Nan Fang Yi Ke Da Xue Xue Bao | 152 | Nanfang yi ke da xue xue bao bian ji bu | China | 1673-4254 | Chinese |
| Curr. Pharm. Des. | 151 | Bentham Science Publishers | Netherlands | 1873-4286 | English |
| Am Fam Physician | 150 | American Academy of General Practice | United States | 1532-0650 | English |
| Pak J Pharm Sci | 138 | Faculty of Pharmacy, University of Karachi | Pakistan | 1011-601X | English |
| Zhongguo Shi Yan Xue Ye Xue Za Zhi | 136 | Zhongguo shi yan xue za zhi she | China | 1009-2137 | Chinese, English |
| Oncology (Williston Park, N.Y.) | 132 | Williston Park | United States | 0890-9091 | English |
| Ying Yong Sheng Tai Xue Bao | 120 | Ying yong sheng tai xue bao bian ji wei yuan hui | China | 1001-9332 | Chinese |
| Zh Nevrol Psikhiatr Im S S Korsakova | 119 | Meditsina | Russia (Federation) | 1997-7298 | Russian |
| J Med Assoc Thai | 116 | Medical Association Of Thailand | Thailand | 0125-2208 | English, Thai |
| Gan To Kagaku Ryoho | 113 | Gan to Kagaku Ryōhōsha | Japan | 0385-0684 | Japanese |

Discussion

To our knowledge, there are no similar studies examining the availability of DOIs in publications archived by PubMed, and such a study of a geographical analysis of countries and continents of publishers has never been done. PubMed was chosen because, contrary to other databases (Web of Science, Scopus), the country of publisher can be easily identified because of a specific field (called Country of Publication). Scopus did not provide this information, even when using the downloading format “all available information”. The Web of Science provides a specific field called “Publisher Address” containing the complete address of the publisher, implying manual or automated isolation of the country from this complete address, which was not only impossible to carry out given the number of articles analyzed in the study, but could be a source of error.

Only one study, published in 2015 (Gorraiz et al. 2016), has assessed the availability of DOIs in Web of Science Core Collection and Scopus over the period 2005-2014. This work showed that, in 2014, the percentage of citable documents (articles, reviews and proceeding papers) with DOIs

in Scopus was about 86% in the health sciences, and 89% in life science. Our results are quite similar: 86.47% of articles had DOIs in PubMed in 2014.

The retrospective analysis of availability of DOIs over the period 1996-2015 has shown that the percentage of articles with DOIs began to increase for articles published in and around the year 2000, with spectacular growth in the years 2002-2003, corresponding quite logically to the years following the launch of DOIs by the IDF. This showed that the DOI system was massively adopted by publishers in the years 2002-2003, meaning that most publishers rapidly understood the numerous advantages related to this system. It is noticeable that the highest percentages of articles with DOIs are mainly found in countries where large scientific publishers dominate (i.e. the United States, the United Kingdom, the Netherlands) and which have been identified by Vardakas et al. (Vardakas et al. 2015) as the most productive countries of publishers in PubMed over the period 2004-2013. Of note is that unfortunately, overall, little effort has been made to assign DOIs retrospectively to articles published before 2000. This is regrettable because this retrospective assignment of DOIs could permit the enhancement of the value of former articles by facilitating their visibility and accessibility.

The current analysis of articles published in 2015 showed that there were still 13.58% of articles that did not have DOIs, corresponding to 15.22% of journals indexed in PubMed which did not yet use DOIs. Publishers who do not assign DOIs must be made aware of the implications and realize, as some already have, the important role DOIs can play in identifying published articles more easily as well as increasing their journal's impact (Braile 2011). More importantly, some scholars state that, in the near future, the value of a publication will be determined by the number of links to it (Wang 2007). Moreover, all of these articles and journals could be excluded from some scientific production evaluations or altmetric measurements, particularly that of the promising service CrossRef Event Data, whose objectives are to track activity surrounding articles or research data with DOIs from potentially any web source. The geographical analysis of countries and continents of publishers showed a high amount of heterogeneity in countries and continents. The analysis also demonstrated that seven of the 50 most productive countries of publishers assigned less than 10% of the DOIs to articles in 2015 (Israel, Malaysia, Mexico, Russia, Thailand, Ukraine and Serbia). For continents, in 2015 Asia was behind in both the percentage of articles with DOIs and the percentage of journals which had

implemented DOIs. The reason for these observations is difficult to interpret because it is an arduous task to determine whether recommendations for publishers to promote DOIs have been specifically implemented in countries or continents. It seems that few countries have implemented this type of recommendation (Gorraiz et al. 2016). Globally, recommendations to promote DOIs can be found in “Recommended Practices for the Presentation and Identification of E-Journals (PIE-J)” (“RP-16-2013 PIE-J (short URL) - National Information Standards Organization” 2013) published in 2013 by the National Information Standards Organization (NISO). Furthermore, the Policy Recommendations for the Open Access to Research Data in Europe (RECODE) project have also published such recommendations: “Publishers should require that data accompanying their publications are citable, and provide clear guidelines for data citation. Data citation should include DOIs” (“RECODE” 2015). As suggested by Park et al. (Park et al. 2011), the further promotion of the core features of the DOI systems, perhaps through case examples and direction from industry/governmental leaders, are essential supporting factors in influencing the promotion of the DOI system's success. Industries and governments that develop DOI-compatible systems are encouraged to promote its features, which should be a powerful influence among organizations that are considering adopting DOI systems. This should have a significant influence on the countries identified in this study as assigning few DOIs, and may convince them to establish recommendations to encourage their publishers to consider using DOIs.

Additionally, the list of the 20 most productive journals lacking DOIs for articles published in 2015 (current analysis), suggests that small and local publishers are mainly implicated. The cost of DOIs may explain the unwillingness of small publishers to implement them.

Nevertheless, publisher fees are proportional to the total publishing revenue, and the deposit fees per DOI (from 0.06 to 1.00 US dollar per content type) do not seem, in any case, to be an impenetrable barrier to the adoption of DOIs (“crossref.org : : publisher fees” 2016). Technical requirements to implement DOIs may also hinder small publishers who work in a traditional way. Nevertheless, it is difficult to believe that in some countries, as in Russia, the 55 of 59 journals that have not implemented DOIs are small publishers, and that the cost and technical requirements could explain this finding. Only a qualitative survey among publishers in different countries would be able to clarify these results. The present study should also encourage publishers who have not yet assigned DOIs to change their strategy in order to benefit from the

numerous advantages offered by the system, as specified in the present article. They should understand that with the new developments surrounding DOIs they may be increasingly on the fringes of scientific publication.

For publishers that have already adopted DOIs for their articles, a way to improve the efficiency of their utilization may be to assign several DOIs referring to different parts of articles (e.g. tables, figures, videos, audio clips, datasets and supporting information), thereby allowing their access as separate entities (see as examples articles published by PLOS One) (DeRisi et al. 2003). As in the case of CrossRef membership publishers ("DOI display guidelines" 2016), suggesting researchers to routinely mention DOIs in references in their instructions to authors would improve researchers' level of awareness of DOIs, and would also allow the readership to move from one article to another at the citation level, regardless of journal publisher. Without full-text citation linking, the user who discovers a desired resource while reading usually has to switch to a different search interface to locate and ultimately access that resource. With DOIs, it only takes a click or two to get to the full text, either as an authorized user or through pay-per-view services ("crossref.org" 2016).

Researchers should be aware that choosing journals that assign DOIs is important, enabling their scientific communications to be more visible and accessible, and ensuring that they can benefit from new metrics such as altmetrics. Moreover, it increases the likelihood that their work will be read and cited.

The results presented in this article pertain only to biomedicine articles from PubMed and includes only articles (the study of Gorraiz et al. (Gorraiz et al. 2016) studied three types of publications: all document types from journals, proceedings series and conferences, and books extracted from Scopus and Web of Science Core Collection). These authors have shown that the social sciences and humanities have lower percentages of documents with DOIs compared to health, life, and physical sciences. It would be useful to complete their study by carrying out a geographical analysis as done in this study, particularly in the social sciences and humanities where practices and issues are different. Furthermore, as some studies have shown that publishing articles in open access journals increases their visibility and therefore increases the number of times they are cited in other articles (Swan 2010), future studies should explore whether there is a relationship between the assignment of DOIs and the number of citations received by articles.

The advantages of developing standards for services are central, DOIs do not waver from this rule. Hence, the increasing use of DOIs should result in the multiplication of useful applications for all those involved in scientific research. In the future, DOIs should become increasingly helpful tools. For this to continue, efforts must be made to generalize their use by publishers, without exception, and in all countries and continents. This should have the consequence of familiarizing researchers with DOIs by means of articles they read, which will allow, in the near future, the development and extension of the use of DOIs to research data. However, errors in assigning DOIs, notably incorrect assignment of a single DOI to multiple articles by bibliometric databases, which exist today (Franceschini et al. 2014), should be limited.

References

- ALM. (2016). Retrieved from <http://alm.plos.org/>
- Altmetric. (2016). Retrieved from <https://www.altmetric.com/>
- Boudry, C., Denion, E., Mortemousque, B., & Mouriaux, F. (2016). Trends and topics in eye disease research in PubMed from 2010 to 2014. *PeerJ*, 4, e1557. doi:10.7717/peerj.1557
- Boudry, C., & Mouriaux, F. (2015). Eye neoplasms research: a bibliometric analysis from 1966 to 2012. *European Journal of Ophthalmology*, 25(4), 357–365. doi:10.5301/ejo.5000556
- Braile, D. M. (2011). After the Impact Factor, the DOI. *Brazilian Journal of Cardiovascular Surgery*, 26(3), I–II. doi:10.5935/1678-9741.20110001
- Brase, J., Lautenschlager, M., & Sens, I. (2015). The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite. *D-Lib Magazine*, 21(1/2). doi:10.1045/january2015-brase
- Carnevale, R., & Aronsky, D. (2007). The life and death of URLs in five biomedical informatics journals. *International Journal of Medical Informatics*, 76(4), 269–273. doi:10.1016/j.ijmedinf.2005.12.001

Chandrakar, R. (2006). Digital object identifier system: an overview. *The Electronic Library*, 24(4), 445–452. doi:10.1108/02640470610689151

Crossref Event Data. (2016). Retrieved from <http://eventdata.crossref.org/>

Crossref initiatives will support reporting to funders | Research Information. (2016). Retrieved from <https://www.researchinformation.info/news/analysis-opinion/crossref-initiatives-will-support-reporting-funders>

crossref.org. (2016). Retrieved from <http://www.crossref.org/>

crossref.org : : publisher fees. (2016). Retrieved from http://www.crossref.org/02publishers/20pub_fees.html

CrossRef's DOI Event Tracker Pilot – Crossref Blog. (2015). Retrieved from <http://blog.crossref.org/2015/03/crossrefs-doi-event-tracker-pilot.html>

DeRisi, S., Kennison, R., & Twyman, N. (2003). The What and Whys of DOIs. *PLOS Biol*, 1(2), e57. doi:10.1371/journal.pbio.0000057

Digital Object Identifier System Handbook. (2016). Retrieved from <http://www.doi.org/hb.html>

DOI display guidelines. (2016). Retrieved from http://www.crossref.org/02publishers/doi_display_guidelines.html

Ducut, E., Liu, F., & Fontelo, P. (2008). An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation. *BMC Medical Informatics and Decision Making*, 8(1). doi:10.1186/1472-6947-8-23

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 22(2), 338–342. doi:10.1096/fj.07-9492LSF

- Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2014). Errors in DOI indexing by bibliometric databases. *Scientometrics*, 102(3), 2181–2186. doi:10.1007/s11192-014-1503-4
- Galligan, F., & Dyas-Correia, S. (2013). Altmetrics: Rethinking the Way We Measure. *Serials Review*, 39(1), 56–61. doi:10.1080/00987913.2013.10765486
- González-Valiente, C. L., Pacheco-Mendoza, J., & Arencibia-Jorge, R. (2016). A review of altmetrics as an emerging discipline for research evaluation. *Learned Publishing*, 29(4), 229–238. doi:10.1002/leap.1043
- 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. doi:10.1016/j.joi.2015.11.008
- 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. doi:10.1371/journal.pone.0120495
- Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 101(2), 1145–1163. doi:10.1007/s11192-013-1221-3
- Home - NLM Catalog - NCBI. (2016). Retrieved from <https://www.ncbi.nlm.nih.gov/nlmcatalog/>
- Honor, L. B., Haselgrove, C., Frazier, J. A., & Kennedy, D. N. (2016). Data Citation in Neuroimaging: Proposed Best Practices for Data Identification and Attribution. *Frontiers in Neuroinformatics*, 10, 34. doi:10.3339/fninf.2016.00034
- ISO 26324:2012 - Information and documentation -- Digital object identifier system. (2012). Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=43506

- Khedmatgozar, H. R., & Alipour-Hafezi, M. (2015). A Basic Comparative Framework for Evaluation of Digital Identifier Systems. *Journal of Digital Information Management*, 13(3), 191.
- Neumann, J., & Brase, J. (2014). DataCite and DOI names for research data. *Journal of Computer-Aided Molecular Design*, 28(10), 1035–1041. doi:10.1007/s10822-014-9776-5
- Park, S., Zo, H., Ciganek, A. P., & Lim, G. G. (2011). Examining success factors in the adoption of digital object identifier systems. *Electronic Commerce Research and Applications*, 10(6), 626–636.
- Paskin, N. (1999). The digital object identifier system: digital technology meets content management. *Interlending & Document Supply*, 27(1), 13–16.
doi:10.1108/02641619910255829
- Paskin, N. (2010). Digital object identifier (DOI) system. *Encyclopedia of Library and Information Sciences*, 3, 1586–1592.
- Plumx. (2015). Retrieved from <https://plu.mx/>
- Rasmussen, P. G., & Andersen, J. P. (2013). Altmetrics: an alternate perspective on research evaluation. *ScieCom Info*, 96(2). Retrieved from
<http://journals.lub.lu.se/index.php/sciecominfo/article/view/7292>
- RECODE. (2015). Retrieved June 13, 2016, from
<http://policy.recodeproject.eu/publishers/policy-content/>
- Rosenblatt, B. (1997). The Digital Object Identifier: Solving the Dilemma of Copyright Protection Online. *Journal of Electronic Publishing*, 3(2).
doi:<http://dx.doi.org/10.3998/3336451.0003.204>
- RP-16-2013 PIE-J (short URL) - National Information Standards Organization. (2013). Retrieved from <http://www.niso.org/publications/rp/rp-16-2013/>

- Sidman, D., & Davidson, T. (2001). A practical guide to automating the digital supply chain with the Digital Object Identifier (DOI). *Publishing Research Quarterly*, 17(2), 9–23.
doi:10.1007/s12109-001-0019-y
- Sieck, S. (2003). *Using the DOI to Improve Profitability In Publishers' E-Commerce Operations*. EPS White Paper Series.
- Simmonds, A. W. (1999). The Digital Object Identifier (DOI). *Publishing Research Quarterly*, 15(2), 10–13. doi:10.1007/s12109-999-0022-2
- Sud, P., & Thelwall, M. (2013). Evaluating altmetrics. *Scientometrics*, 98(2), 1131–1143.
doi:10.1007/s11192-013-1117-2
- Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Retrieved from <http://eprints.ecs.soton.ac.uk/18516>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLOS One*, 8(5), e64841.
doi:10.1371/journal.pone.0064841
- Tolwinski, A. (2015). Need to Know What's Going On with an Article DOI? The Wait Is Over. *Science Editor*, 38(3/4), 106–108.
- United Nations Statistics Division- Standard Country and Area Codes Classifications (M49). (2016). Retrieved from <http://unstats.un.org/unsd/methods/m49/m49regin.htm>
- Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics*, 9(3), 570–576. doi:10.1016/j.joi.2015.05.002
- Vardakas, K. Z., Tsopanakis, G., Pouloupoulou, A., & Falagas, M. E. (2015). An analysis of factors contributing to PubMed's growth. *Journal of Informetrics*, 9(3), 592–617.

- Wagner, C., Gebremichael, M. D., Taylor, M. K., & Soltys, M. J. (2009). Disappearing act: decay of uniform resource locators in health care management journals. *Journal of the Medical Library Association : JMLA*, 97(2), 122–130. doi:10.3163/1536-5050.97.2.009
- Wang, J. (2007). Digital object identifiers and their use in libraries. *Serials Review*, 33(3), 161–164. doi:10.1016/j.serrev.2007.05.006
- Zotero | Home. (2016). Retrieved from <https://www.zotero.org/>