

Article

Research Articles about Open Access Indexed by Scopus: A Content Analysis

Rosângela Schwarz Rodrigues ^{1,*}, Vitor Taga ^{1,*} and Mariana Faustino dos Passos ²

¹ Graduate Program of Library and Information Science, Federal University of Santa Catarina, Florianópolis 88.040-900, Brazil

² Library Studies, Federal University of Santa Catarina, Florianópolis 88.040-900, Brazil; maripassos23@gmail.com

* Correspondence: rosangela.rodrigues@ufsc.br (R.S.R.); vitor.taga@gmail.com (V.T.)

Academic Editor: Barbara Meyers Ford

Received: 31 May 2016; Accepted: 22 September 2016; Published: 11 October 2016

Abstract: This study analyzes research articles about open access (OA) indexed by the Scopus database, published from 2001 to 2015, in order to: (a) propose a categorization scheme about OA; (b) categorize the scientific production about OA; and (c) identify research trends on OA through disciplines at international level over time. The authors used descriptive statistical methods and deductive content analysis using an unconstrained matrix in 347 selected research articles. The most explored themes were found to be “overview, current state, and growth of OA” counting for 98 articles (28.2%), and “awareness, perceptions, and attitudes toward OA” for 75 articles (21.6%). As a conclusion, this study reveals a continuous and growing research interest by the OA community in studies focused on case studies regarding the development or evolution of OA in relation to certain groups, institutions, regions, periods, and how different actors perceive and address the OA movement.

Keywords: open access; scientific production about open access; research articles about open access; content analysis

1. Introduction

In most areas of knowledge, the scientific journal is established as a source of reference for formal scientific communication [1–3], thus the peer-reviewed journal article has become the most accepted type or support for formal communication of scientific research results [1–6], recognized as a widely accepted formal mechanism of certification, dissemination, and scientific knowledge memory [3]. Ziman [7] considered the article published in specialized journals as the most important means of scientific communication.

Changes in information and communication technologies have opened space for questioning the prevailing costs and value added by commercial publishers in the scientific editorial system. The concentration of titles by a few publishers is one of the reasons for that, as only three major commercial publishers (Elsevier, Springer-Kluwer, and Wiley-Blackwell) hold 42% of all published articles and the most prestigious and the largest circulation journals [8]. A further 2000 publishers hold the other titles, none with more than 3% of the total. Larivière, Haustein, and Mongeon ([9], p. 9) argue that “the top commercial publishers have benefited from the digital era, as it led to a dramatic increase in the share of scientific literature they published. It has also led to a greater dependence by the scientific community on these publishers.” In addition to digital factors, the strategies of merging and concentration of titles practiced by the big commercial editors raise concerns in relation to antitrust regulatory policies [10] and boycotts from librarians, governments, and researchers.

The “Saint Matthew effect” (or accumulated advantage) generated from the domination of journal titles with higher prestige [11] propel a cycle in which the core journals remain forever on top, and other titles cannot break the barrier and achieve top visibility.

In the history of scientific communication, publication of articles in scholarly journals has begun to be used as a regulatory resource [12] and conclusion mechanism [2] of scientific activities, being used as a criterion for academic and scientific career promotion [13,14], an indicator of competence and professional recognition, as well as attested quality research [1].

Scientific journal publication is controlled by a small group of large commercial publishers that offer their Big Deals to university libraries, which are now found with increasingly reduced or unchanged funds. Because of the journal crisis, in the late twentieth century and early twenty-first, several initiatives emerged, including the open access movement [1,4].

According to Weller [15], publication in open access began in the 1990s, from the change caused by the publication nature of digital content disseminated on the Internet, inspired by the open source communities. Laakso ([16], p. 23) considers that “the birth and growth of OA (open access) can be attributed to a variety of interrelated technology, ideology, policy, economic, and legal factors.” The most important arguments include (a) Information technology advanced both in processing capacity and in reach, going from local to globally interconnectivity. Perfect copies of documents could be distributed at close to zero marginal cost; (b) The questioning of the old practice of academic institutions paying escalating subscription fees in order to gain access to research they participated in producing; (c) Academic faculty and researchers author, review, and act as editors for journal article manuscripts; (d) Libraries have been reported to struggle with the equation of managing a severely constrained budget while facing rising subscription expenses; (e) Research is, to a large extent, directly and indirectly funded by taxpayer funds and should be considered a public good; (f) Academic authors write for impact, not for royalties, thus, it is not in the authors’ interest to restrict access and readership of their written work.

According to Suber [17] and Eve [18], by making research results widely available and useful, open access concurrently benefits readers, authors and the advance of research itself. Open Access helps readers in finding, retrieving, reading, and using the research they need, while favoring authors to expand their audience, reaching readers who may apply, cite and build their research based on them, and thus expand impact and, as a result, advance research and all consequent benefits. For Suber [17], open access brings benefits not only to scientific research, but to all.

Gold open access refers to research freely available online in its original form by the journal that published it [16,18], regardless of the journal’s business model [17] and licenses for using the article, so that any reader can have online access to the full content of the published paper without cost [15]. Regarding the journals’ business model, gold open access can be characterized into five types: *full journal immediate open access*, *platinum open access*, *hybrid open access*, *delayed open access* and *promotional open access* [15,16].

In the *full journal immediate open access* model, the journal publishes articles and other content in open access, these journals can operate with or without the charge of Article Processing Charge (APC) to authors who wish to publish research [15,16]. Journals operating in this business model that do not charge APC are entitled *platinum open access*, usually these journals are supported by scientific societies or universities, where the dissemination of research has priority over the financial gain [15].

The model named *hybrid open access* refers to journals with paid subscribed access that offer authors the option to individually publish the article in open access in the journal by paying APC [15–18]. This type of journal should not even be named open access, since most papers are restricted, moreover, most journals charge full subscription, even when OA to part of the content has already been paid by the author. Similarly, the so-called *delayed open access*, which could not be named “open” because it does not allow access to articles when researchers need. The *promotional open access* refers to subscribed access journals that offer individual articles or entire issues in open access sporadically until further notice or temporarily with a certain expiration time [16].

Green Open Access refers to the indirect free access to some version of the article that was self-archived elsewhere than the journal's website that published it, regardless of their business model and their permissions to use the article. This deposit is usually performed by the author on a personal website, blog, institutional or thematic repository, and social media, but an embargo period may be stipulated by the publisher, so the self-archiving will only be authorized after that period [15,17]. The so-called green route has not changed the publishing landscape of scientific publications.

According to Weller ([15], p. 49) "the publication in open access is no longer a minority purpose, reserved to those with particular concern about it, it has become a dominant practice" that, according to Eve [18], despite being a worldwide phenomenon of international roots and implementations, particularly in South America, recent controversies and disputes regarding the transition and normalization of scientific publications for open access have occurred in the Anglophone academic setting, due to the urgency in its implementation deriving from strong open access recommendations adopted by research funders.

Weller [15] argues that the Finch Report, the most important document of a country on access, has been criticized for its negligence to academic interests. The document also displays the strong influence of commercial publishers in establishing recommendations to protect their interests and reinforce their power. Such recommendations serve to support a self-perpetuating elite which may increase the already fierce competition for research funding, in addition to doubling the generated cost because the subscriptions to these journals will most likely not be canceled.

The theme OA is a relevant subject to the entire scientific community interested in the effects emerging from free access to knowledge usage, therefore researchers from different disciplines have studied and published about it since day one. For these reasons, this research analyzes the publication themes of research journal articles about open access (OA) indexed by the Scopus database, published from 2001 to 2015, in order to:

- (a) Propose a categorization scheme about OA;
- (b) Categorize the scientific production about OA; and
- (c) Identify the research trends on OA on the international scene over time.

Related Studies

The earliest study addressing OA themes identified in the literature was *The Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals*, published in 2005 by Charles W. Bailey Jr. [19], which presents over 1300 resources regarding OA including research studies published in peer-reviewed journals from 1999 to 2004. Each one was classified in different categories created by the author under the Budapest Open Access Initiative definition.

In 2006, M. Carl Drott [20] conducted an *Open Access* review published in the *Annual Review of Information Science and Technology* (ARIST). He attempted to set out the main discussions regarding OA publishing through several sources including research studies published in peer-reviewed journals. In this review, the author organized his discussion in different topics, which indicated and explored the main themes about OA at that moment, even though it was not his main intention.

The first identified study that explicitly focused on investigating the main OA themes was published in 2007 by Zao Liu and Gang Wan [21] in the *Chinese Librarianship* journal. Using content analysis, they developed an OA classification scheme and identified the publication themes and trends in scholarly journal articles on open access in library and information science literature from 2000 to 2005.

In 2010, Charles W. Bailey Jr. [22] published *Transforming Scholarly Publishing through Open Access: A Bibliography*, another compilation containing primarily books and journal articles published from 1999 to 2010. Although the main purpose of this investigation was not to identify and discuss the main themes about OA, its listing, as well as his first bibliography, established a perspective of the main topics regarding OA.

A few years later, in 2014, Rongying Zhao e Shengnan Wu [23] published an article in the journal *Scientometrics*; analyzing the main OA research themes in academic journals published in China as well as exploring their development situation and status by conducting a co-word and social network analysis.

In 2016, Sandra Miguel, Ely F. T. de Oliveira and Maria C. C. Grácio [24] published an article in the journal *Publications*, analyzing the worldwide scientific production on OA. The authors identified the main sub-themes pertaining to the theme OA through co-occurrence of words in the titles, abstracts, and keywords of 1179 articles.

Although the research regarding OA themes has been published, with the exception of the classification scheme proposed by Liu and Wan [21] based on scholarly journal articles in library and information science literature from 2000 to 2005, no other classification scheme has been suggested or updates made.

2. Data and Methods

This study is a descriptive and exploratory, bibliographic and documentary research [25], and regarding the problematic approach, it can be classified as quantitative and qualitative [25–28], quantitative because a descriptive statistical method is used, and qualitative as content analysis is also used.

Content analysis is a versatile, unobtrusive, pragmatic, context-sensitive research method that uses a set of procedures for subjective analysis of data, based on objective, systematic, and quantitative description of the manifest content of communication or specified characteristics of message [29–39].

With a deductive approach, content analysis starts with a theory or relevant findings as guidance for initial codes [33], the categories are established prior to the analysis, and once the categories are agreed on, a categorization matrix is developed creating the initial coding scheme [32–35]. A coding scheme is a translation instrument that organizes data into categories guiding coders to make decisions in the content analysis [32–35].

The purpose of creating categories is to provide a means of describing the phenomenon, to increase understanding and generate knowledge [30]. When using an unconstrained matrix, different categories are created within its bounds, following the principles of inductive content analysis ([32], p. 111).

In an inductive content analysis, also described as conventional content analysis or emergent content analysis, coding categories are derived directly from the text data, moving from the specific to the general, so that particular instances are observed and then combined into a large whole or general statement. Usually, inductive content analysis is used with a study design whose aim is to describe a phenomenon, in cases where no previous studies dealing with it, when existing theory or research literature is limited or when researchers instead of using preconceived categories, choose to immerse themselves in the data to allow categories and their names to flow from data and new insights to emerge [32,33].

The *corpus* of this study was composed of articles about open access published in scientific journals indexed by the Scopus database, because it has twice as many titles and over 50% more publishers listed than any other multidisciplinary database, with over 21,500 peer-reviewed journals. Thus, Scopus is considered the largest multidisciplinary database of specialized scientific literature [40].

The year 2001 is the starting point of our sample because this was the year of the Budapest Open Access Initiative (BOAI), an event in December 2001, with the publication of the Budapest Declaration, which formalized a first definition of open access [41]. The year 2015 was established as the end of the period, assuming that the articles published that year would already be indexed in Scopus in the period in which this study was conducted (March and April 2016).

We included articles of research results only, written in English, Spanish, and Portuguese. Publications such as rapid communications, short communication, technical notes, letters to the

editor, books, article reviews, viewpoints, editorials, and calendar events, were excluded. Our criteria is similar to those used by Grandbois and Beheshti [42], Liu and Wan [21], and Togia and Korobili [43].

According to Öchsner ([44], p. 11), a research result paper refers to an “original full-length manuscript which has not been published previously, except in a preliminary form” and that “describes a highly significant advancement in the particular field of research.” These papers are “judged according to originality, novelty, quality of scientific content, and contribution to existing knowledge. They include full introduction, methods, results, and discussion sections.” In addition, according to Liu and Wan [21], journal paper studies that use the content analysis method often limit their scope to full-length articles.

Only the articles containing the terms “open access” or “open-access” which addressed themes in line with the open access movement were included. Other papers about OA used similar criteria, such as Bailey Jr. [19,22]; Grandbois and Beheshti [42]; Liu and Wan [21]; Miguel, Oliveira and Grácio [24]; and Zhao and Wu [23].

In March 2016, a search for the terms “open access” and “open-access” in quotation marks was carried out in the field *title* in Scopus, and only journal articles published from 2001 to 2015, written in English, Spanish, and Portuguese were selected, retrieving 1261 articles. This result was based on the titles provided by the proxy of the Federal University of Santa Catarina (UFSC).

The examination of titles, abstracts, keywords, references, and authors of the articles was done in order to ensure that the final *corpus* would include only articles relevant to the study. To select research results articles, we identified the ones containing the sections: methods, results, and references. After the selection, 482 articles that did not address issues relevant to the open access movement were discarded. Of the remaining 779 relevant articles, however, 432 did not meet the inclusion criteria. Of these, 360 articles did not present research results, 19 articles were written in other languages than the selected ones, four articles were repeated, one article did not have the term “open access” in the title, one article had a broken link, and 47 articles did not provide access to the full text. After all these eliminations 347 articles were included in the final sample.

The classification scheme on OA proposed by Liu and Wan [21] was adopted, in order to categorize the 347 articles about OA selected by this study. In order to analyze the publication trends of 227 scholarly journal articles on open access in library and information science literature from 2000 to 2005, Liu e Wan [21] used the method of content analysis according to a classification scheme developed by them, which covers five broad categories regarding the types of: (a) journal; (b) article; (c) author; (d) country; and (e) content. In respect to the classification scheme concerning the content type, by drawing on the existing literature on open access at that time, the authors established nine sub-categories and classified each article into one of them. These sub-categories are: (1) *General works* (general discussion of open access, open access journals, open access archives; open access initiatives/organizations; literature reviews/bibliographies); (2) *Stakeholders* (open access and roles and perspectives of authors, researchers, libraries/librarians, general public, publishers/editors, universities, and professional associations); (3) *Legal issues* (copyright of open access journals, open access archives, or both; legislations concerning open access); (4) *Economic issues* (business models of open access journals, open access archives, or both); (5) *Technical issues* (software, models, designs, protocols for open access journals, open access archives, or both); (6) *Quality control* (peer review of open access journal articles, self-archiving articles, or both); (7) *Research impact* (impact and citations of open access journal articles, self-archiving articles, or both); (8) *Specific open access cases* (planning and implementation of open access journals, open access archives, or both); and (9) *Open access and developing countries* (open access movement and developing countries; open access journals and developing countries; open access archives and developing countries).

After conducting a preliminary examination of articles published from 2013 to 2015, we found that the categories were not mutually exclusive and exhaustive. For example, the category “*General Works*” seems a miscellaneous or residual category, despite the fact that this sort of category is usually dedicated to units that occur rarely or are uncodable for other reasons. However, it does

not appear to be the case. The category “open access and developing countries” generalizes studies related to developing or emerging countries, classifying them into one single category, which ends up disregarding the results and contributions presented in these studies, and addresses the issue of open access as an exclusive problem to developing countries and the considered central ones. In addition, some of the topics addressed in the preliminarily analyzed articles were not clearly represented in any of the categories, for example, studies investigating ethical issues in open access publications, use of social media for online sharing of scientific papers in open access, perception of students in relation to open access, planning and implementation of open access policies in universities and countries, among others.

In relation to Bailey Jr. open access bibliographies [19,22], based on the Budapest Open Access Initiative (BOAI), in 2005, the author classified 1300 resources regarding OA, including research papers, published from 1999 to 2004. In 2010, the author updated his first bibliography, this time, focusing mainly on books and journal articles published from 1999 to 2010, a total of 1100 references were covered in his second bibliography. In both works, by listing the works and resources regarding open access, the author clustered the main topics emerging from the resources and/or references analyzed, thus, he was able to classify each one into categories and sub-categories developed by the author himself, some of them more general and some more specific. The main categories identified in both bibliographies were: (a) *General Works*; (b) *Copyright Arrangements for Self-Archiving and Use*; (c) *Open Access Journals* (focusing more on economic issues); (d) *E-prints*; (e) *Disciplinary Archives*; (f) *Institutional Repositories*; (g) *Open Archives Initiative and OAI-PMH*; (h) *Conventional Publishers Perspectives*; and (i) *Open Access Legislation, Government Reviews, Funding Agency Mandates, and Policies*; and (j) *Open Access in Countries with Emerging and Developing Economies*.

By comparing the categories and sub-categories created by Bailey Jr. [19,22] with the classification scheme concerning the content type developed by Liu and Wan [21], similarities are perceived. Similar patterns were also perceived with the main topics covered by Drott [20] in his review of open access, and, despite being somewhat less similar, with the research themes found by Zhao and Wu [23], and the main sub-themes registered by Miguel, Oliveira and Grácio [24].

Drott [20] organized his review in sections, each one exploring a main topic regarding open access, which were: (a) *Overview*; (b) *Financial Pressures on Libraries*; (c) *Journal Quality and Reputation*; (d) *Peer Review—Quality Control*; (e) *Editorial processing*; (f) *Archiving*; (g) *Archiving as an Alternative to Open Access Journals*; (h) *Author Fees for Open Access*; (i) *Concerns About Open Access*; (j) *Finding Archived Material*; (k) *Copyright and Ownership*; (l) *Organizations Supporting Open Access*; and (m) *Developing Trends*.

Zhao and Wu [23], by conducting an author-keywords analysis coupling relationship and clustering those high-frequency keywords, were capable of identifying seven research themes regarding open access in China, namely: (a) OA of the government’s information resource; (b) Influence of OA over the information sharing and scholarly communication; (c) OA journal and OA repositories; (d) Quality evaluation of OA journals; (e) Development strategy of OA; (f) OA Publishing of academic journals; and (g) Institutional repositories’ building strategy

Likewise, Miguel, Oliveira and Grácio [24], from the co-occurrences of keywords registered in the articles analyzed, was able to highlight the main sub-themes regarding open access, namely: (a) dissemination and information access; (b) articles; (c) journals; (d) repositories; (e) libraries; and (f) bibliometrics.

Considering the similarities between the main topics, (sub)themes and/or (sub)categories regarding open access, the coding scheme proposed by Liu and Wan [21] was adopted as a guideline for our initial codes. However, we followed the principles of inductive content analysis which recommends coding categories to be derived directly from the text data during inductive content analysis, moving from specific to general, so that particular instances are observed and then combined into a large whole or general statement [32,33].

Once the final sample of 347 articles was determined and the preliminary examination was carried out using the classification scheme by Liu and Wan [21], it was decided to create a printed file for each article containing a numerical identification followed with the title of the article in bold and a descriptive sample extracted from the article which indicated the main purpose or aim of the research presented, in most cases, the original description of the purpose or aim was used.

As this open coding process continued, the labels were coded into sheets on IBM SPSS Statistics 20 software, categories were generated at this stage and grouped by contrasting those considered similar or dissimilar into broader higher order categories in a tree-shape diagram, reducing their numbers and leading to our initial coding scheme.

In order to reliably describe all aspects of the contents, two researchers of the team were assigned as coders to conduct the content analysis independently. During their readings, individual notes and labels were made for the purpose of capturing first impressions and key thoughts. To increase the reliability of the categorization research, citations were also used to point out from where or from what kinds of original data categories were formulated. After that, the coding scheme was developed and the process of categorization started. Once completed, before reporting our findings, we measured the reliability by calculating the percentage of agreement between the two raters using Cohen's Kappa coefficient. Kappa is calculated as follows:

$$K = \frac{P_A - P_C}{1 - P_C}$$

where: P_A is the proportion of units on which the raters agree, and P_C is the proportion of units for which agreement is expected by chance.

Cohen's Kappa coefficient measures the agreement between two raters, "approaches 1 as coding is perfectly reliable and goes to 0 when there is no agreement other than what would be expected by chance" ([34], p. 5).

3. Results and Discussions

3.1. Categorization Scheme

By conducting a deductive content analysis using an unconstrained matrix, eight main themes were identified in the articles, which became the main categories in our version of the classification scheme proposed by Liu and Wan [21].

Table 1 presents the adapted classification scheme. For each category, individual codes, descriptions, subcategories, and citations were provided, in this case, a research articles' purpose or aim was cited, as they clarify the main theme of each article.

Table 1. Categorization Scheme.

Code	Categories	Descriptions	Subcategories	Examples
Perception	Awareness, perceptions and attitudes toward Open Access (OA)	Scientific research aimed to explore underlying attitudes, experiences, perceptions, reactions, opinions, engagements, views, awareness, and/or other stakeholders' behaviors with OA.	OA initiatives; movement; values; philosophy; practices; journals; articles; publishing; self-archiving; repositories; policies; principles; models; literature; thesis and dissertations; resources; and other issues.	<i>This study explores PhD faculty members' current awareness of open access (OA) and perceptions of OA publishing, focusing on demographic characteristics to understand whether these variables correspond to specific perceptions and behaviors.</i>
Growth	Overview, current state and growth of OA	Scientific research aimed to provide an overall growth and/or current scenario of OA.	OA movement; initiatives; e-books; theses and dissertations; sources; journals; repositories; publishing; articles; policies; indicators; self-archiving; publications; literature; open archives.	<i>The main aim of this study was to obtain a holistic view of the current status of OA journals in China.</i>
Impact	OA citation performance and other impact measures	Scientific research aimed at measuring the impact of open access (OA) in terms of citations' number, performance and other impact measures like readership, downloads, access, and/or impact factor.	Literature; publishing; publications; journals; articles; books; documents; articles versions.	<i>The aim of this study was to compare OA and subscription journals in terms of the average number of citations received both at the journal and article level.</i>
Economics	OA economics and its implications on the publishing market	Scientific research aimed at analyzing the economics of OA.	OA business models; sustainability; viability; article processing charges; funding sources; interlibrary loan; publication fees; journal costs; subscription prices; publishers' profits and other economic issues.	<i>In this article we regard the implications of different OA models for scholars, publishers, libraries, and funding organizations and try to explain the motivations behind the actions currently taking place on the scientific publishing market.</i>
Technology	Technical development, system features and other technological issues	Scientific research aimed to focus on technical development and technological innovations in OA.	Metadata; OAI-PMH; Web 2.0; usability; web references; software; interoperability; system features and other technical and technological issues.	<i>The purpose of this paper is to evaluate the interoperability of ten open access repositories in the field of computer science and IT.</i>
Quality	Quality control and visibility	Scientific research aimed to verify quality and/or visibility mostly regarding the indexing of OA journals and peer-review process.	Publishing practices; web presence; quality criteria; ranking; peer review system; versions; articles; manuscripts; accuracy; transparency; indexing.	<i>The present study makes an attempt to measure the effect of peer review process on research impact of publications in comparison to those which have not gone through the process via citation analysis.</i>
Legal	Legal and ethical aspects	Scientific research aimed to investigate legal and/or ethical aspects related to OA.	OA policies; provisions; statements; clauses; licensing; ethical and legal regulations and procedures; legislations; copyright; patent; intellectual property rights and other legal and ethical issues.	<i>This paper reviews the efforts of several organizations to implement open access clauses within the context of their content licenses and discusses some of the challenges inherent in this activity.</i>
Philosophy	OA movement, philosophy, values and principles	Scientific research aimed to discuss broadly the OA movement, its principles, values, and philosophy.	Scientific communication; production; information; knowledge; articles; journals; development.	<i>The paper seeks to reconsider open access and its relation to issues of "development" by highlighting the ties the open access movement has with the hegemonic discourse of development and to question some of the assumptions about and scientific communication upon which the open access debates are based.</i>

3.2. Agreement Matrix

The proportion of agreement between raters’ judgments about the classification of articles is presented in Table 2. The horizontal cells (rows) represent the judgments of the first rater and the vertical cells (columns) represent the judgments of the second rater. The proportion of articles classified in the same categories is presented diagonally drawn on the table (from top left to bottom right), highlighted in light gray, which added up represent the total proportion as observed (P_A) by both raters.

Table 2. Inter-rater agreement matrix.

		Rater2								Total
		Count (Expected Count)								
		Perception	Growth	Impact	Economics	Legal	Quality	Technology	Philosophy	
Rater1	Perception	70 (16.2)	0 (18.6)	1 (10.8)	1 (9.9)	0 (4.1)	0 (5.2)	2 (8.0)	1 (2.2)	75
	Growth	1 (21.6)	83 (24.8)	3 (14.4)	4 (13.3)	0 (5.5)	1 (6.9)	3 (10.7)	5 (2.9)	100
	Impact	0 (8.9)	0 (10.2)	40 (5.9)	1 (5.4)	0 (2.2)	0 (2.8)	0 (4.4)	0 (1.2)	41
	Economics	2 (9.9)	2 (11.4)	3 (6.6)	38 (6.1)	0 (2.5)	0 (3.2)	1 (4.9)	0 (1.3)	46
	Legal	0 (4.3)	0 (5.0)	1 (2.9)	0 (2.7)	19 (1.1)	0 (1.4)	0 (2.1)	0 (0.6)	20
	Quality	1 (5.8)	0 (6.7)	1 (3.9)	0 (3.6)	0 (1.5)	23 (1.9)	1 (2.9)	1 (0.8)	27
	Technology	1 (7.6)	1 (8.7)	1 (5.0)	2 (4.6)	0 (1.9)	0 (2.4)	30 (3.7)	0 (1.0)	35
	Philosophy	0 (0.6)	0 (0.7)	0 (0.4)	0 (0.4)	0 (0.2)	0 (0.2)	0 (0.3)	3 (0.1)	3
	Total	75	86	50	46	19	24	37	10	347

When the values in the diagonally drawn cells were added (highlighted in light gray), we obtained a total of 306 articles classified in the same categories by both raters (P_A), representing 88.18% of the 347 analyzed articles. From the sum of the multiplications between the row total by the column total corresponding to each agreement cell and dividing this product by the total observations, we obtained a total of 59.74 frequencies expected by chance (P_C).

When we performed Kappa calculations, we obtained an agreement measure (K) of 0.857, with an Asymptotic Standard Error of 0.021 (not assuming the null hypothesis), an Approx. T of 36.523 (using the asymptotic standard error assuming the null hypothesis), and an Approx. Sig. of 0.000, which shows a good estimate, with low variability and significant agreement degree in the population, as shown in the sample.

In order to maintain a consistent nomenclature to describe the relative strength of agreement associated with Kappa statistics, categorical labels were assigned to the corresponding ranges of Kappa, using as a benchmark the classification proposed by Landis and Koch [45], that establishes six categories assigned to corresponding ranges of Kappa, as shown in the following Table 3.

Table 3. Agreement interpretation proposed by Landis and Koch ([45], p. 165).

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

According to the benchmarks suggested by Landis and Koch [45] for interpreting Kappa coefficient strength of agreement, our study’s coefficient is classified as “almost perfect agreement”, which means that the categorization has reached a good overall agreement, also satisfactory

trustworthy results. Nevertheless, the judgments in disagreement, total of 41 (11.82%) were evaluated by the lead investigator of this study for tie-breaking purposes, so that the other analyses proposed in this study were not restricted to articles that had their classification agreed between the two raters. The next section presents the frequency of articles by category containing the tiebreaker judgments.

3.3. Categorization of the Scientific Production about Open Access

By conducting a descriptive statistic to verify the frequency and percentage distribution of research articles per category, shown in Table 4, this study was able to disclose the most prominent categories regarding articles.

Table 4. Frequency and percentage distribution of articles about OA between categories.

Categories	Frequency	Percent	Cumulative Percentage
Growth	98	28.2%	49.9%
Perception	75	21.6%	21.6%
Economics	46	13.3%	75.2%
Impact	42	12.1%	62.0%
Technology	36	10.4%	98.6%
Quality	26	7.5%	88.2%
Legal	19	5.5%	80.7%
Philosophy	5	1.4%	100%
Total	347	100%	

In descending order, the most common themes were “Overview, current state and growth of OA” with 98 articles (28.2%), “Awareness, perceptions and attitudes toward OA” with 75 articles (21.6%), “OA economics and its implications on the publishing market” with 46 articles (13.3%), “OA citation performance and other impact measures” with 42 articles (12.1%), “Technical development, system features and other technological issues” with 36 articles (10.4%), “Quality control and visibility” with 26 articles (7.5%), “Legal and ethical aspects” with 19 articles (5.5%), and “OA movement philosophy, values and principles” with five articles (1.4%).

This result demonstrated that most articles aimed to provide an overall growth or current scenario of OA in relation to stakeholders, institutions, regions, time, and/or disciplines. Articles regarding OA movement, philosophy, values, and principles were poorly represented, being the less exploited theme, probably because this type of content is commonly explored in non-research result papers, without empirical data, which were not considered for this study. In the second position were papers related to perceptions, often authors’ perceptions about OA, considering issues such as journal rankings and intentions to publish in open access journals.

3.4. Trends in OA Themes

Figure 1 introduces the trends in the evolution of OA themes over time, it shows the number or frequency of articles for each category per year. Although 2001 had been stated as the initial year for the period proposed for this study, no research result article about OA was registered until 2003.

Through the chronological perspective shown in Figure 1, it was possible to verify that in the first three years (2003, 2004, and 2005) no specific category was prominent. In 2003 only one research result article was registered, related to the theme “Legal and ethical aspects”; in the following year, 2004, two research articles were identified, one concerning the theme “Overview, current state and growth of OA” and the second one on the theme “OA economics and its implications on the publishing market”. In 2005, three research result articles were reported, one focused on the theme “Awareness, perceptions and attitudes toward OA”, the second on the theme “OA citation performance and other impact measures” and the third one regarding the theme “Quality control and visibility”.

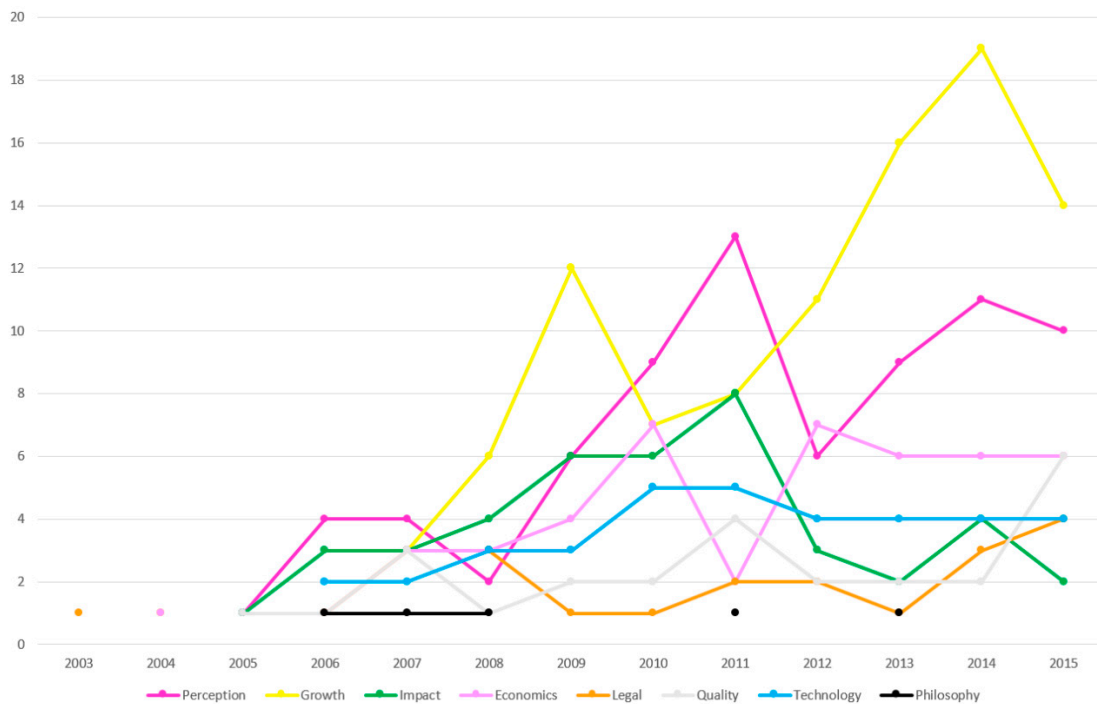


Figure 1. Evolution of the Open Access Themes in the International Scientific Literature.

From a general perspective is clear that the number of research result articles has increased, especially from 2006, when all the eight themes counted at least one publication and a few had already demonstrated some increase. This observation was also identified by Miguel, Oliveira, and Grácio ([24], p. 6) when the authors presented the evolution of the theme OA in Scopus, starting from 1982 to 2014. In the same way, our results indicated a meaningful starting point in terms of number of publications about OA from 2005 to 2007, growing in the following years with one small decrease in 2012. Since our research followed until 2015, we also observed another decrease in 2015.

Regarding 2012, our results demonstrated that only two themes presented an increase concerning published articles: “OA economics and its implications on the publishing market” with seven articles, five more than in the previous year; and “Overview, current state and growth of OA” with 11 articles, three more than in the previous year.

Concerning the year 2015, two themes presented growth of articles, two remained with the same number of articles in comparison with the previous year, three presented a small fall, and one did not present articles in that year. The two themes that presented growth were “Legal and ethical aspects” and “Quality control and visibility”.

The two themes that remained with the same number of papers were “Technical development, system features and other technological issues” and “OA economics and its implications on the publishing market”; the three themes that presented a small decrease in the number of articles were “Awareness, perceptions, and attitudes toward OA”, “Overview, current state, and growth of OA” and “OA citation performance and other impact measures”.

It is important to note that these three themes presented an increase from the previous year, so this type of decrease can be expected to occur occasionally. The only theme that did not have any articles published in 2015 was “OA movement, philosophy, values, and principles”.

Considering the themes individually, only the theme “OA movement, philosophy, values, and principles” did not demonstrate a continuous growth across the years and the theme “OA citation performance and other impact measures” indicated a decrease in number of the articles published after 2011, probably because this type of research became secondary in the investigations, serving as supplementary empirical data.

The two most prominent themes in number of articles published were “Awareness, perceptions, and attitudes toward OA” and “Overview, current state, and growth of OA”, the other themes presented significant growth but less significant in comparison with these two. These results reveal a continuous and growing research interest by the OA community in studies focused on case studies regarding the development or evolution of OA in relation to certain groups, institutions, regions, and periods and how different actors perceive and address the OA movement.

4. Conclusions

The content categorization of Open Access papers shows a diversity of approaches, however, it is still a challenge to identify classifications able to be recognized as valid and ubiquitous by all investigations. Articles originate from different countries yet have similar concerns. However we did not identify any special treatment based on geographic region, which leads us to conclude that OA is addressed from a global perspective, considering that the issue of access to scientific literature in the digital era is everyone’s concern. As we were able to clearly identify the evolution of the literature due to the creation of the term ‘open access’ identifying all the publications, we suggest a categorization of the research themes, which identify more clearly and accurately the studies on the theme and allow the identification of gaps and uncovered themes. Our categorizations result in eight themes: Growth, Perception, Economics, Impact, Technology, Quality, Legal, and Philosophy.

The results of our categorization of the articles analyzed shows the tendency of researchers’ major concern to be analyzing the movement itself (Growth with 98 articles, 28.2%) and the researchers’ perceptions (with 75 articles, 21.6%), in a sort of self-reflection and search for affirmation of the movement. If we consider the 360 articles excluded from the sample because they did not present research results, it is possible to deduce a major number of papers published in relevant journals discussing the theme OA with a more reflexive approach, which is a questionable choice, since the arguments are economic: high priced access impacts all researchers. The economic aspects represent just 13% of the articles, with 46 articles, followed by Impact with 42 papers (12%). These results show the concerns with the costs involved in the editorial activity, which is a key element in all discussions and the basic objective of researchers: to be cited by their peers as effectively contributing to the area. According to the arguments found in the literature review, those are the most relevant themes to be investigated.

It is also noteworthy that no changes were verified in papers using the methods as the ones in this sample in the last years—either the discussions are stable or this represents the need to change research choices to bring new contributions to the movement.

Author Contributions: Vitor Taga and Rosângela Schwarz Rodrigues conceived, designed, analyzed the data and wrote the paper Mariana Faustino dos Passos analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. United Nations Educational, Scientific and Cultural Organization (UNESCO). *Introduction to Open Access*; UNESCO: Paris, France, 2015.
2. Fjällbrant, N. Scholarly communication—Historical development and new possibilities. In Proceedings of the International Association of Technology University Libraries Conference, Trondheim, Norway, 1997. Available online: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1389&context=iatul> (accessed on 30 May 2016).
3. Owen, J.S.M. The Scientific article in the age of digitization. Ph.D. Thesis, Universiteit van Amsterdam, Amsterdam, The Netherlands, November 2005.
4. United Nations Educational, Scientific and Cultural Organization (UNESCO). *Scholarly Communications: Open Access for Researchers*; UNESCO: Paris, France, 2015.
5. Tenopir, C.; King, D.W. *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*; Special Libraries Association: Washington, DC, USA, 2000.

6. Tenopir, C.; King, D.W. A importância dos periódicos para o trabalho científico. *Rev. Bibliotecon. Bras.* **2001**, *25*, 15–26.
7. Ziman, J. *The Force of Knowledge: The Scientific Dimension of Society*; Cambridge University Press: London, UK, 1977.
8. McGuigan, G.S.; Russell, R.D. The business of academic publishing: A strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing. *Electron. J. Acad. Spec. Librariansh.* **2008**, *9*, 1–11.
9. Larivière, V.; Haustein, S.; Mongeo, P. The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE* **2015**, *10*, 1–15. [[CrossRef](#)] [[PubMed](#)]
10. McCabe, M. The Impact of Publisher Mergers on Journal Prices: Theory and Evidence. *Ser. Libr.* **2001**, *40*, 157–166. [[CrossRef](#)]
11. Merton, R.K. *The Sociology of Science: Theoretical and Empirical Investigations*; University of Chicago Press: Chicago, IL, USA, 1973.
12. Frohmann, B. The role of the scientific paper in science information systems. *J. Educ. Libr. Inf. Sci.* **2000**, *42*, 13–28.
13. Meadows, A.J. *Communicating Research*; Academic Press: London, UK, 1998.
14. Ziman, J. *Public Knowledge: The Social Dimension of Science*; Cambridge University Press: London, UK, 1968.
15. Weller, M. *The Battle for Open: How Openness Won and Why It Doesn't Feel Like Victory*; Ubiquity Press: London, UK, 2014.
16. Laakso, M. *Measuring Open Access: Studies of Web-Enabled Innovation in Scientific Journal Publishing*; Edita Prima: Helsinki, Finland, 2014.
17. Suber, P. *Open Access*; The MIT Press: Cambridge, UK, 2012.
18. Eve, M.P. *Open Access and the Humanities: Contexts, Controversies and the Future*; Cambridge University Press: Cambridge, UK, 2014.
19. Bailey, C.W., Jr. *Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals*; Association of Research Libraries: Washington, DC, USA, 2005.
20. Drott, M.C. Open Access. *Annu. Rev. Inf. Sci. Technol.* **2006**, *40*, 79–109. [[CrossRef](#)]
21. Liu, Z.; Wan, G. Scholarly Journal Articles on Open Access in LIS Literature: A Content Analysis. *Chin. Librariansh.* **2007**, *23*, 1–11.
22. Bailey, C.W., Jr. *Transforming Scholarly Publishing through Open Access: A Bibliography*; Digital Scholarship: Houston, TX, USA, 2010.
23. Zhao, R.; Wu, S. Study on themes and authors' influence of open access in China. *Scientometrics* **2014**, *101*, 1165–1177. [[CrossRef](#)]
24. Miguel, S.; Oliveira, E.F.T.; Grácio, M.C.C. Scientific Production on Open Access: A Worldwide Bibliometric Analysis in the Academic and Scientific Context. *Publications* **2016**, *4*, 1–15. [[CrossRef](#)]
25. Creswell, J.W. *Projeto de Pesquisa: Métodos Qualitativos, Quantitativo e Misto*, 3rd ed.; ARTMED: Porto Alegre, Brazil, 2010.
26. Connaway, L.S.; Powell, R.R. *Basic Research Methods for Librarians*, 5th ed.; Libraries Unlimited: Santa Barbara, CA, USA, 2010.
27. Vaughan, L. *Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using and Interpreting Statistics*, 4th ed.; Information Today Inc.: Medford, NJ, USA, 2008.
28. Goldenberg, M. *A Arte de Pesquisar: Como Fazer Pesquisa Qualitativa em Ciências Sociais*; Record: Rio de Janeiro, Brazil, 2007.
29. Berelson, B. *Content Analysis in Communication Research*; Free Press: New York, NY, USA, 1952.
30. Cavanagh, S. Content analysis: Concepts, methods and applications. *Nurse Res.* **1997**, *4*, 5–13. [[CrossRef](#)] [[PubMed](#)]
31. Downe-Wamboldt, B. Content analysis: Method, applications and issues. *Health Care Women Int.* **1992**, *13*, 313–321. [[CrossRef](#)] [[PubMed](#)]
32. Elo, S.; Kyngäs, H. The qualitative content analysis process. *J. Adv. Nurs.* **2007**, *62*, 107–115. [[CrossRef](#)] [[PubMed](#)]
33. Hsieh, H.; Shannon, S.E. Three Approaches to Qualitative Content Analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288. [[CrossRef](#)] [[PubMed](#)]
34. Stemler, S. An Overview of Content Analysis. *Pract. Assess. Res. Eval.* **2001**, *7*, 1–10.

35. Weber, R.P. *Basic Content Analysis*; Sage Publications: London, UK, 1985.
36. Prasad, B.D. Content Analysis: A method in Social Science Research. In *Research Methods for Social Work*; Lal Das, D.K., Bhaskaran, V., Eds.; Rawat: New Delhi, India, 2008; pp. 173–193.
37. Rosengren, K.E. (Ed.) *Advances in Content Analysis*; Sage Publications: London, UK, 1981.
38. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*; Sage Publications: London, UK, 1980.
39. Krippendorff, K. Content Analysis. In *International Encyclopedia of Communication*; Barnouw, E., Gerbner, G., Schramm, W., Worth, T.L., Gross, L., Eds.; Oxford University Press: New York, NY, USA, 1989; pp. 403–407.
40. Elsevier. Scopus: Content. Available online: <https://www.elsevier.com/solutions/scopus/content> (accessed on 30 May 2016).
41. BOAI. Ten Years on from the Budapest Open Access Initiative: Setting the Default to Open. Available online: <http://www.budapestopenaccessinitiative.org/boai-10-recommendations> (accessed on 30 May 2016).
42. Grandbois, J.; Beheshti, J. A bibliometric study of scholarly articles published by library and information science authors about open access. *Inf. Res.* **2014**, *19*, 1–19.
43. Togia, A.; Korobili, S. Attitudes towards open access: A meta-synthesis of the empirical literature. *Inf. Serv. Use* **2014**, *34*, 221–231.
44. Öchsner, A. *Introduction to Scientific Publishing: Backgrounds, Concepts, Strategies*; Springer: New York, NY, USA, 2013.
45. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).