



Universiteit
Leiden
The Netherlands

The deliverance of open access books : examining usage and dissemination

Snijder, R.

Citation

Snijder, R. (2019, January 29). *The deliverance of open access books : examining usage and dissemination*. Retrieved from <https://hdl.handle.net/1887/68465>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/68465>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/68465> holds various files of this Leiden University dissertation.

Author: Snijder, R.

Title: The deliverance of open access books : examining usage and dissemination

Issue Date: 2019-01-29

5 Better sharing through licenses? : Measuring the influence of Creative Commons licenses on the usage of open access monographs

Snijder, R. (2015). Better Sharing Through Licenses? Measuring the Influence of Creative Commons Licenses on the Usage of Open Access Monographs. *Journal of Librarianship and Scholarly Communication*, 3(1), eP1187. <https://doi.org/10.7710/2162-3309.1187>

5.1 Introduction

Open access (OA) and content licenses are closely intertwined. The first Budapest open access Initiative declaration (Chan *et al.*, 2002) – widely seen as the official birth of the open access movement – does not explicitly state the need for a license, but the Berlin Declaration (“Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities,” 2003) and the Bethesda Statement (Suber *et al.*, 2003) pose two conditions: a license permitting distribution and reuse; and a deposit of the contents in a technically suitable manner. The goal of the open access movement is to disseminate scholarly and scientific knowledge as widely as possible, and using licenses to systematically remove the barriers created by copyright restrictions is an important tool.

One of the best-known licenses used to achieve this is the Creative Commons (CC) license. The Creative Commons organisation describes its set of licenses as a “simple, standardized way to give the public permission to share and use your creative work – on conditions of your choice.” (“About - Creative Commons,” n.d.) These licenses enable the copyright owner to allow certain types of usage – such as copying or modifying the contents – while constricting other forms of use, for instance by prohibiting commercial reuse. The six licenses vary in the amount of restrictions placed on the reuse of the work.

Beyond alerting individual users of their reuse rights, there is another important aspect to these licenses. Placing Creative Commons license code on digital content not only provides a human readable license, but also

provides a machine readable version of the license, enabling computers to determine in what way the content may be reused (Lessig, 2004). Machine readable licenses enable others to create automated services, based on the type of reuse granted by the content owner.

Although both Creative Commons and the open access movement seek to restore the balance between the owners of creative works and the prospective users, not all Creative Commons licenses are considered equally 'open' by OA proponents. For example, the BOAI now recommends a specific CC license: CC-BY (Budapest Open Access Initiative, 2012). According to the open access Scholarly Publishers Association (OASPA) this license allows for unrestricted reuse of content, as long as the source work is appropriately attributed (Redhead, 2012).

The preference for this license is not undisputed¹³; and has led to discussions about the relative merits of the degree of openness provided by the different CC licenses. In this paper, we will use a simpler distinction: documents which are available without charge, and documents that are not only available without charge, but also made available under a license that enables reuse and further dissemination. Peter (Suber, 2008) coined the terms 'gratis' and 'libre' OA to distinguish between these two distinct forms of 'open access'. Throughout this paper, books published under *any* type of CC license are categorised as *libre* open access; all other freely accessible books are categorised as *gratis* OA. In other words, books published under open licences ranging from CC-BY to CC-BY-NC-ND are here defined as *libre*;¹⁴ books which are only 'free to read' and may not be freely used otherwise are defined as *gratis*.

Because documents which have been released under a *libre* license – such as Creative Commons – carry fewer barriers to reuse, it stands to reason that such content is easier to share and more likely to be used. This paper examines a discrete collection of open access monographs – the OAPEN Library collection – in an attempt to determine whether *libre* licenses do in fact lead to greater use of open access works.

13 A recent example is the interview with Paul Royster (Poynder, 2014).

14 Although the most restrictive CC licenses do not permit the adaptation of content, they still allow greater reuse than a *gratis* license that restricts users to the personal use rights under copyright law.

5.2 The OAPEN Library and the DOAB

The OAPEN Library was officially launched in September 2010 (OAPEN Consortium, 2011). It is a web-based collection of open access monographs, published by dozens of publishers. In December 2013, the collection contained over 2,000 titles by 55 publishers. The OAPEN Library offers several ways to make its contents accessible: it enables searching and browsing, readers can share book descriptions via social media, and it also offers several data feeds (Open Access Publishing in European Networks, 2010b). In the OAPEN Library, books are made available under several licenses: roughly 50% of the titles are disseminated under a Creative Commons license, while the rest are made available under a more restrictive regime. In other words, about half the titles in the OAPEN Library are available as *gratis* OA, the other half as *libre* OA.

The OAPEN Library is managed by the OAPEN Foundation. In April 2012, the OAPEN Foundation launched the Directory of Open Access Books (DOAB) as a discovery service for open access books (“DOAB: Directory of Open Access Books,” n.d.). The directory is open to all academic publishers and aims to contain as many books as possible, provided that these books are peer-reviewed and published under an open access license. In addition to the publishers already taking part in OAPEN, several other academic publishers have placed their books in the DOAB (Snijder, 2013a). The DOAB is connected to the OAPEN Library, and automatically uploads descriptions of new titles from OAPEN. However, not all books in the OAPEN Library are listed in the DOAB: it only contains the titles with a Creative Commons – or a comparable – license. The selection is not limited to CC-BY, but extended to the full range of CC licenses. So, while the OAPEN Library is a mixture of *gratis* and *libre* OA, the DOAB only lists *libre* OA books. This automated aggregation – based on the machine-readable licence information – results in extra exposure of the *libre* books in the OAPEN Library.

5.3 Examining the Impact of Licenses on use

The OAPEN Library and the DOAB are examples of two types of use of open access works: use by individual end users and use by intermediaries, who provide additional services built on or around open content. Here, the end users are the readers of the books contained in the OAPEN Library. Use by this group can be measured by counting the number of times a book has been downloaded from the OAPEN Library. Downloads have been

chosen as a metric for two reasons. First, readers both within and beyond universities are able to download books held in the OAPEN Library. This contrasts to approaches to measuring impact that are based on counting citations, which only capture a specific form of use by academic readers. Second, downloads can be measured directly from the OAPEN server. This ensures a fast, and dependable, result. Although it is not possible to equate a download with further use (e.g. reading, integration into other work), we can assume that a much-downloaded online monograph has been read more often than a book which has been downloaded just a few times. We cannot, however, state that 100 downloads equate to 100 people reading the book cover to cover.

While all books in the OAPEN Library are openly available to download by end users, a significant proportion are also available under *libre* open access licenses. These types of licenses allow intermediaries like the DOAB to aggregate books and display them on a website, which creates another access point for individual users. This type of aggregation would not be possible with books available under a more restrictive *gratis* license.

In this paper, we examine both types of use – by individual end users and by intermediaries – as we consider the effects of *gratis* and *libre* licenses on the number of times books in the OAPEN Library are downloaded. In looking at this, we will make two comparisons: first, between *gratis* and *libre* books that were available in the OAPEN Library prior to the creation of the Directory of Open Access Books; and, second, between *gratis* books only available in the OAPEN Library and *libre* books available in the OAPEN Library and also included in the DOAB. This will allow us to measure whether *libre* books are downloaded more often in general, as well as whether additional aggregation has a significant effect on downloads. Our guiding research question for this study is:

Compared to *gratis* access, does applying an open license (*libre* access) have a positive effect on the number of times an open access book is downloaded?

Although the question of license effect is of primary interest, we are aware that the use of books in the OAPEN Library may also be influenced by factors other than license type, such as the subject or the language of the monographs. Earlier research published by (Snijder, 2013b) described the differences in number of downloads per subject in 2011. It seems reasonable to assume that subject still affects downloads. We could also argue that the language of a publication acts as a barrier to use: when readers cannot understand the language, the books become useless to them. And because

of the length of the texts involved, the chances of successful automatic translation are slim.

The number of times a book is downloaded might reasonably be expected to reflect the size of a particular language community. Therefore, it is important to remain alert to the impact of both subject and language on use when attempting to understand this ecosystem. Regardless of the type of license applied to a work, and whether or not it is made available in an aggregation service like the DOAB, prospective readers are not very likely to download books on subjects that are of no interest to them, or written in languages they cannot read.

Finally, it may be useful to note that information about the license for each individual book is always available to the users of the OAPEN Library website: each page describing a monograph contains a description of the license. Moreover, information about the complete collection can be obtained through several data feeds. On top of a description of the books, all data feeds also list the license information. Within the OAPEN Library, there is no technical distinction between books with a *libre* license or a more restrictive license: each monograph can be searched and downloaded in the same manner. Differences in usage cannot, therefore, be accounted for by restrictions in the infrastructure.

5.4 Literature review

There are three areas of literature relevant to this study: the conflicting interests of creators and users; the use of Creative Commons licenses to balance these interests, and the impact of Creative Commons licenses on usage.

5.4.1 Tensions between the interests of creators and users

Much of the debate around intellectual property, particularly copyright, centres on the tensions between creators' rights and users' rights. A much cited example is the paper by (Landes & Posner, 1989), in which they discuss the optimal level of copyright protection. This entails balancing the interests of the creators of a work versus the people who want to use it – either as a 'reader' or for creating a derivative work. The conflicting interest of these stakeholders is also described by (Boldrin & Levine, 2002). In their analysis, intellectual property law has two components. The first component is the right to own ideas and sell them. The second component

is the right to control the use of those ideas after sale. They argue that the second component – termed “downstream licensing” – leads to monopolies, impairing economic welfare. Again, we see the need to balance the interests.

Rather than finding a balance in current copyright law, (Suzor, 2014) argues that in certain cases, a high level of copyright protection is not needed. According to Suzor, many content users are prepared to pay the producers, even if the content is freely available. Choosing an intellectual property model that allows free use, while encouraging – but not enforcing – financial support may both enhance dissemination and compensate producers.

5.4.2 Balancing interests using Creative Commons licenses

Several authors have discussed the legal context around Creative Commons Licenses. Loren (2007) criticizes the “climate of overly broad ownership rights for creative works”, and argues that it hinders the use and reuse of creative works. The complexity of the current copyright system leads to high costs, which disadvantages individuals who do not have the same financial resources as corporations. Broadly applying Creative Commons licenses helps to create a “semicommons of creative works” (Loren, 2007, pag. 328), which enables a greater and more diverse usage – to the benefit of society. This argument closely resembles the removal of legal barriers in the Berlin Declaration and the Bethesda Statement, describing a right to access and reuse scholarly and scientific content.

Hietanen (2008) also describes the advent of Creative Commons licenses as a reaction to the way copyright law has developed. Hietanen discusses the implications of applying CC licenses in great detail and analyses the license-choosing process and the clauses of the Creative Commons licenses. The approach by Kim (2007) is slightly different, and tries to understand the motives of CC licensors through surveys and interviews. Again, the conflict of interests of the different stakeholders are debated. However, Kim’s paper attempts to categorize the types of content licensed under Creative Commons, and the motives of the content owners. The paper describes a large variety of content types. Furthermore, the reasons to use a CC license vary: some content owners place emphasis on the public benefits, while others are motivated by more personal reasons. (Morrison, 2012) discusses the application of CC licenses within open access publishing. According to the author, the goals of OA publishing and CC licenses are not aligned. She concludes that the lack of restrictions of the CC-BY license actually might be harmful to OA; the absence of restrictions leaves the author or content

owner without tools to control its reuse – suggesting that some licenses may tip the balance too far.

5.4.3 Do Creative Commons licenses enhance usage?

Despite Morrison's (2012) concerns, other authors arrive at more optimistic conclusions regarding Creative Commons licenses. (Carroll, 2006) looks at CC licenses and the changing role of intermediaries. The licenses are made machine readable, which opens new possibilities for those who enable all kinds of transactions based on the licensed works. The image sharing website Flickr.com is a well-known example: it enables end users to find pictures published under licenses that allow reuse. (Guibault, 2011) discusses the relation between authors of scientific and scholarly works and copyright ownership in the European context. She concludes that licensing documents under Creative Commons (partly) increases access and reuse.

There is little to no research published on the effects of *gratis* versus *libre* open access, especially in the realm of monographs. (J. L. Hilton III, Lutz, & Wiley, 2012) investigated revisions made to academic textbooks published under an open license. They conclude that – in line with expectations – the amount of revisions is relatively low. This is consistent with the findings in this paper; an open license does not automatically lead to a surge in usage. As far as could be established, there is no literature available which aims to quantify variances in usage based on differences in licenses.

5.5 Methods and the data set

The OAPEN Library platform logs usage data, starting from January 2011. Among the data recorded is the number of times each monograph has been downloaded in a month. We will use this as an indicator of successful dissemination: more downloads means a better result. For this paper, we will analyse the data captured over a period of 33 months: from January 2011 up until September 2013. During this time, 1,734 different books were made available through the OAPEN Library. Of these monographs, 855 were disseminated as *libre* open access and 879 were distributed under a more restrictive regime. Of the 855 *libre* titles, 512 were published under a CC-BY-NC-ND licence; the most restrictive open license. In contrast, only 4 titles were available under the CC-BY license. The rest of the titles were licensed as follows: 162 titles under CC-BY-ND; 168 titles under CC-BY-NC; and 9 titles under CC-BY-SA.

The Directory of Open Access Books was launched in April 2012, 16 months after January 2011. To understand whether the DOAB influences usage, we will compare the data of the first 15 months of the OAPEN Library to the data of the following 18 months. In the first 15 months, a total of 935 monographs were disseminated via the OAPEN Library; 563 of those under a *libre* license. After that period, the collection grew to the 1,734 books, as described above. The monthly download data for each book is used; if a book has been available for 33 months, this leads to 33 samples. Of course, not all books were available during that period, but the total number of samples used in our analysis is over 34,000.

Table 1 lists the number of books that were made available in the OAPEN Library, split by period and license. In the period before the launch of the Directory of Open Access Books, the difference in usage is not very large: on average, books published under a *libre* license were downloaded 29 times per month, compared to 21 times per month for books with *gratis* licenses. However, in the period after the DOAB launch, the difference widens to 84 downloads versus 34 downloads on average per month. It seems reasonable to assume that the aggregation in the DOAB has a positive influence.

Table 1 Average downloads per period

| Period | Libre OA | | Gratis OA | |
|---|-----------------|---------------------|-----------------|---------------------|
| | Number of books | Mean downloads (SD) | Number of books | Mean downloads (SD) |
| Direct use only (Jan. 2011-Mar. 2012) | 563 | 29.6 (66.0) | 372 | 21.9 (37.4) |
| Aggregation and/or direct use (Apr. 2012-Sep. 2013) | 855 | 84.1 (409.1) | 879 | 34.5 (44.7) |

The OAPEN Library contains books on many subjects; our dataset contains 96 different subject classifications. Nevertheless, not all subjects are equally spread among the collection. Among the most common subjects we find Politics & Government and History. When looking at the licenses used, it becomes clear that they are not evenly spread: for instance, 22% of the books on Politics & government are published under a *libre* license, compared to 61% of books on History. Table 2 contains a more comprehensive listing.

Table 2 Subjects in the OAPEN Library

| Subject (BIC classification) | Total number of books | Percentage | Books: <i>libre</i> license | Percentage (of all books) | Books: <i>gratis</i> license | Percentage (of all books) |
|---------------------------------|-----------------------|------------|-----------------------------|---------------------------|------------------------------|---------------------------|
| Politics & government (JP) | 398 | 23.0% | 93 | 5.4% | 305 | 17.6% |
| History (HB) | 237 | 13.7% | 144 | 8.3% | 93 | 5.4% |
| Society & culture: general (JF) | 129 | 7.4% | 75 | 4.3% | 54 | 3.1% |
| Economics (KC) | 107 | 6.2% | 14 | 0.8% | 93 | 5.4% |
| Sociology & anthropology (JH) | 77 | 4.4% | 24 | 1.4% | 53 | 3.1% |
| Other subjects | 786 | 45.3% | 505 | 29.1% | 281 | 16.2% |
| Total | 1,734 | 100% | 855 | 49.3% | 879 | 50.7% |

The collection contains monographs in several languages. Most are written in English, Dutch, German or Italian, but also books in Danish, Latin or Russian are made available. As is the case with subject, the portion of books published under a *libre* license varies strongly per language; while 57% of books in English can be downloaded using a *libre* license, the percentage for Dutch is much lower: 13%. Table 3 lists the number of books per language.

Table 3 Languages in the OAPEN Library

| Language | Total number of books | Percentage | Books: <i>libre</i> license | Percentage (of all books) | Books: <i>gratis</i> license | Percentage (of all books) |
|-----------------|-----------------------|------------|-----------------------------|---------------------------|------------------------------|---------------------------|
| English | 711 | 41.0% | 408 | 23.5% | 303 | 17.5% |
| Dutch | 494 | 28.5% | 62 | 3.6% | 432 | 24.9% |
| German | 346 | 20.0% | 303 | 17.5% | 43 | 2.5% |
| Italian | 118 | 6.8% | 74 | 4.3% | 44 | 2.5% |
| Other languages | 65 | 3.7% | 8 | 0.5% | 57 | 3.3% |
| Total | 1,734 | 100% | 855 | 49.3% | 879 | 50.7% |

The complete data used for this paper is available at <http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-8ut1-25>.

5.6 Analysis

Our analysis starts with measuring the effect of four factors – license, DOAB aggregation, subject, and language – on usage. This helps to determine if all factors indeed affect the number of downloads in the OAPEN Library. If one or more of them is not relevant, it can be discarded from our analysis. The one-way independent ANOVA statistical method is used to check whether each influence has a statistically significant effect. This procedure tests if the differences between the mean downloads of the books can be explained by chance. The results of each individual test are summarized in Table 4.

Table 4 Effects of the factors

| Influence | Results |
|---|---|
| License | There was a significant effect of license on monograph downloads, $F(1, 19575.517) = 195.114$, $p < .001$, $\omega^2 = 0.00$ |
| DOAB aggregation | There was a significant effect of DOAB aggregation on monograph downloads, $F(1, 25226.413) = 277.956$, $p < .001$, $\omega^2 = 0.00$ |
| Subject | There was a significant effect of subject on monograph downloads, $F(10, 5995.946) = 46.935$, $p < .001$, $\omega^2 = 0.00$ |
| Language | There was a significant effect of language on monograph downloads, $F(4, 10528.836) = 248.871$, $p < .001$, $\omega^2 = 0.01$ |
| Results: The assumption of homogeneity of variance has been violated; therefore, the Welch F-ratio is reported. | |

When a statistically significant effect has been measured, the differences between the analysed groups is bigger than can be expected by chance. However, in large groups this will happen more often; our data set contains over 34,000 samples. The height of the F -ratio indicates the effect size: a higher ratio indicates a stronger effect of the experiment – in our case: license; aggregation through the DOAB; subject and language. Furthermore, the ω^2 value describes the proportion of the variance between the two groups. If the value of ω^2 is 0.01, this means that approximately 1% of the difference in downloads can be attributed to the effect investigated.

The results show that usage of the OAPEN Library is not only influenced by license; it is also affected by DOAB aggregation, subject and language. This complicates the goal of identifying the specific influence of license type. A common way to proceed is to use the multifactor ANOVA procedure to measure the effect of license, combined with the impact of DOAB aggregation, subject and language. Nevertheless, in order to get meaningful results from this procedure, several requirements must be met. The most important precondition is the homogeneity of variance. In other words, the

means used in the procedure should be evenly distributed. Unfortunately, our data does not meet this condition. As a possible solution to overcome the statistical problems, the data is split into smaller subsets.

We have seen before that the usage in the period before the DOAB launch is strongly different from the usage patterns in the period after the launch. To compensate for this, the data is split in two sets: the usage statistics generated in the period before the launch of the directory – January 2011-March 2012 – and the number of downloads registered in the period when DOAB aggregation was deployed – April 2012-September 2013. The groups to be analysed share the same subject or the same language, and the data was gathered in the same period. So, for instance, the usage of all books with the subject Politics & government in the period before the launch of the DOAB – January 2011-March 2012 – is analysed to see whether the license has a significant influence. Splitting up the data creates smaller subsets; but even the smallest group – Sociology & anthropology, in the period January 2011-March 2012 – contains 415 samples.

The analysis focuses on the impact of licensing on direct use (usage pre-DOAB launch), and on the impact of licensing on direct use and aggregated use (usage post-DOAB launch). When we look at use prior to the launch of the DOAB, we expect that simply using *libre* licenses will have a positive effect on the number of times books are downloaded by readers. When we examine use after the launch of the Directory of Open Access Books, the role of a *libre*-enabled intermediary in providing an additional access point is analysed. With regard to aggregators like the DOAB, it is expected that *libre* licenses will enhance the number of downloaded books indirectly--by facilitating additional access points which stimulate readers to find and download books.

As we have seen, the results are not only affected by the license used; the effects of subject and language also play a prominent role. The effects of language and subject are not straightforward: whether or not a certain language or subject enhances or diminishes the number of books downloaded is hard to predict. In contrast, the use of a *libre* license is directly aimed at removing barriers to usage. The impact of subject and language can be seen in the analysis below in that the influence of licenses varies per dataset. However, the overall picture is clear: the use of *libre* licenses alone has a limited impact on downloads, while aggregating *libre*-licensed books has a positive effect on the number of books downloaded.

5.6.1 Impact of licensing on OAPEN downloads

5.6.1.1 Subjects and license

Here, the difference in mean number of downloaded books is examined between *libre* and *gratis* books that share the same subject. This analysis only includes usage prior to the launch of the DOAB. The results are mixed: for the books on History or the books on Society & culture, the license has no effect on the number of books downloaded. However, for the other subsets, the differences in mean number of downloaded books is statistically significant.

Even where there is a significant difference, the effects of publishing under an open license are not very large. Before, we discussed the *F*-ratio and the ω^2 value as an indication of the impact. If we look at these numbers, it becomes clear that the effect of *libre* licenses for books on Economics is much smaller compared to the other subsets. Also, the ω^2 value is never higher than 0.02. In other words: *libre* licenses do not always lead to a difference on the number of books downloaded; when such a difference is found, the influence of licences is much smaller for books on Economics and for other groups the measured impact is no more than approximately 2%.

Table 5 lists the mean number of downloads per subject in the time before the launch of the DOAB.

Table 5 Subjects and license; direct use only

| Subject | <i>Libre</i> license | | <i>Gratis</i> license | | Results |
|---------------------------------|----------------------|------------------|-----------------------|------------------|--|
| | N | Mean downl. (SD) | N | Mean downl. (SD) | |
| Politics & government (JP) | 969 | 26.6 (36.8) | 1169 | 17.3 (22.2) | There was a significant effect of license on monograph downloads, $F(1, 1525.148) = 47.376$, $p < .001$, $\omega^2 = 0.02$ |
| History (HB) | 1136 | 20.6 (22.3) | 785 | 21.5 (28.9) | No significant effect of license on monograph downloads could be found, $F(1, 1919) = 0.7$, $p = .403$, $\omega^2 = 0.00$ |
| Society & culture: general (JF) | 635 | 46.8 (171.1) | 263 | 37.6 (107.4) | No significant effect of license on monograph downloads could be found, $F(1, 896) = 0.655$, $p = .418$, $\omega^2 = 0.00$ |

| Subject | <i>Libre</i> license | | <i>Gratis</i> license | | Results |
|-------------------------------------|----------------------|------------------------------|-----------------------|------------------------------|--|
| | N | Mean downl. (<i>SD</i>) | N | Mean downl. (<i>SD</i>) | |
| Economics (KC) 213 | | 41.4 (81.1) | 569 | 25.8 (31.7) | There was a significant effect of license on monograph downloads, $F(1, 236.640) = 7.446$, $p = .007$, $\omega^2 = 0.00$ |
| Sociology & anthropology (JH) | 356 | 30.1 (34.8) | 59 | 14.5 (13.1) | There was a significant effect of license on monograph downloads, $F(1, 221.856) = 38.333$, $p < .001$, $\omega^2 = 0.00$ |
| Other subjects | 3466 | 29.5 (44.7) | 2116 | 21.7 (30.8) | There was a significant effect of license on monograph downloads, $F(1, 5502.144) = 58.887$, $p < .001$, $\omega^2 = 0.01$ |

Results: With the exception of "History (HB)" and "Society & culture: general (JF)", the assumption of homogeneity of variance has been violated; therefore, the Welch F -ratio is reported.

5.6.1.2 Languages and license

Here we follow the same procedure: the data is split into groups with the same language in order to create groups with equal attributes. The data in Table 6 was captured before launching the DOAB.

Again we see that license type does not create a statistically significant difference in all groups, and that both the F -ratio and the ω^2 value are relatively low in the groups where a statistically significant difference is found. The maximum ω^2 value is even lower compared to the analysis on subject: it is 0.01. In other words, the biggest measured impact of licenses is approximately 1%. Moreover, the books written in Italian and other languages – where no significant statistical differences were found – show a different download pattern: the mean downloads of books with a *libre* license is lower compared to the group of *gratis* titles.

Table 6 Languages and license; direct use only

| Language | Libre license | | Gratis license | | Results |
|-----------------|---------------|------------------|----------------|------------------|---|
| | N | Mean downl. (SD) | N | Mean downl. (SD) | |
| English | 3883 | 35.4 (83.3) | 2233 | 27.6 (47.9) | There was a significant effect of license on monograph downloads, $F(1, 6113.989) = 21.867$, $p < .001$, $\omega^2 = 0.00$ |
| Dutch | 598 | 24.6 (24.9) | 978 | 21.0 (29.0) | There was a significant effect of license on monograph downloads, $F(1, 1574) = 6.074$, $p = .014$, $\omega^2 = 0.00$ |
| German | 1221 | 26.5 (30.8) | 433 | 20.2 (31.1) | There was a significant effect of license on monograph downloads, $F(1, 752.804) = 13.153$, $p < .001$, $\omega^2 = 0.01$ |
| Italian | 1052 | 14.9 (25.1) | 586 | 16.3 (21.7) | No significant effect of license on monograph downloads could be found, $F(1, 1357.292) = 1.382$, $p = .240$, $\omega^2 = 0.00$ |
| Other languages | 21 | 9.0 (9.8) | 731 | 10.9 (12.5) | No significant effect of license on monograph downloads could be found, $F(1, 750) = .492$, $p = .483$, $\omega^2 = 0.00$ |

Results: With the exception of "Dutch" and "Other languages", the assumption of homogeneity of variance has been violated; therefore, the Welch F-ratio is reported.

5.6.1.3 Conclusion on the impact of licenses on downloads

From a statistical point of view, the number of downloaded books is sometimes positively affected by open licenses. However, we have also seen that if there is a positive effect, it is very small. Furthermore, not all groups of books are affected by the license. If the books are grouped by subject, for the titles on History and the books Society & culture – 21% of all titles – the difference in number of books downloaded is not caused by the license. When the books are grouped by language, we see a statistically significant effect for monographs written in English and Dutch – almost 70% of all titles – with an associated ω^2 value of 0.00. An effect of approximately zero percent is not very large.

We can conclude that the impact of *libre* licenses is limited – the download behaviour of users of the OAPEN Library is not affected in any practical way by the type of license used. However, in the next section we will see that *libre*-enabled aggregation through an intermediary has a much bigger effect on usage.

5.6.2 Impact of license-enabled aggregation on OAPEN Downloads

5.6.2.1 Subjects and aggregation

When we look at the download data for the period after the launch of the DOAB, the results are quite different. Compared to their *gratis* counterparts, each group of monographs published under a *libre* license and so listed in the DOAB is downloaded more. Here, the mean number of downloads of books under a *libre* license is almost twice as high compared to *gratis* titles. In the previous data set, the difference is closer to 25%.

In addition, not only are the differences in mean downloads larger, but the statistical effects are also more profound. First, the *F*-ratios – defining the size of the effect we are measuring – are much higher compared to the data set listed in Table 5. Also, the values of ω^2 are much bigger. In the case of Sociology & anthropology it is 0.17; about 17% of the difference could be explained by the *libre* license and the subsequent aggregation through the Directory of Open Access Books. Table 7 lists the data of the monographs grouped by subject.

Table 7 Subjects and license; aggregation and direct use

| Subject | Libre license (Access: OAPEN and DOAB) | | Gratis license (Access: OAPEN only) | | Results |
|-------------------------------|--|---------------------|--|---------------------|---|
| | N | Mean downl. (SD) | N | Mean downl. (SD) | |
| Politics & government (JP) | 1812 | 69.6 (54.4) | 1516 | 34.8 (37.7) | There was a significant effect of license on monograph downloads, $F(1, 3218.685) = 468.751$, $p < .001$, $\omega^2 = 0.12$ |
| History (HB) | 1507 | 88.3 (159.6) | 2894 | 24.3 (28.0) | There was a significant effect of license on monograph downloads, $F(1, 1554.432) = 237.930$, $p < .001$, $\omega^2 = 0.09$ |

| Subject | Libre license (Access: OAPEN and DOAB) | | Gratis license (Access: OAPEN only) | | Results |
|---------------------------------|---|---------------------|--|---------------------|---|
| | N | Mean downl. (SD) | N | Mean downl. (SD) | |
| Society & culture: general (JF) | 1109 | 87.4 (64.4) | 666 | 42.8 (42.6) | There was a significant effect of license on monograph downloads, $F(1, 1756.454) = 306.974$, $p < .001$, $\omega^2 = 0.12$ |
| Economics (KC) | 352 | 99.0 (62.1) | 849 | 39.6 (38.4) | There was a significant effect of license on monograph downloads, $F(1, 466.097) = 276.973$, $p < .001$, $\omega^2 = 0.10$ |
| Sociology & anthropology (JH) | 757 | 73.7 (55.9) | 72 | 35.1 (24.7) | There was a significant effect of license on monograph downloads, $F(1, 153.424) = 117.562$, $p < .001$, $\omega^2 = 0.17$ |
| Other subjects | 6491 | 86.9 (549.6) | 4386 | 38.8 (55.2) | There was a significant effect of license on monograph downloads, $F(1, 6683.120) = 49.062$, $p < .001$, $\omega^2 = 0.00$ |

Results: The assumption of homogeneity of variance has been violated; therefore, the Welch F-ratio is reported.

Languages and aggregation

When the titles are grouped by language, the statistical effects of a *libre* license leading to aggregation by the DOAB are also visible. Most interesting are the differences in F -ratios and ω^2 values between the different language groups. While the *libre* titles written in Dutch and the titles written in “Other languages” clearly benefit from the aggregation, the effects on books in English and German are less noticeable. Still, the findings are statistically significant, and another metric is also clearly pointing in the same direction. If only direct usage is analysed – the data in Table 6 – the difference between mean number of downloads of books on a *gratis* licence is small; the average amount of downloaded *gratis* books is almost as high as the mean number of downloads of books on a *libre* license. However, the data in Table 8 depicts a much larger difference. Here, the mean number of downloads of *libre* books is almost twice the amount for *gratis* books.

Table 8 Languages and license; aggregation and direct use

| Language | Libre license (Access: OAPEN and DOAB) | | Gratis license (Access: OAPEN only) | | Results |
|-----------------|---|------------------|--|------------------|---|
| | N | Mean downl. (SD) | N | Mean downl. (SD) | |
| English | 6245 | 118.3 (565.4) | 4018 | 50.7 (51.7) | There was a significant effect of license on monograph downloads, $F(1, 6406.638) = 88.388$, $p < .001$, $\omega^2 = 0.01$ |
| Dutch | 962 | 55.8 (35.5) | 4031 | 22.9 (38.5) | There was a significant effect of license on monograph downloads, $F(1, 1547.706) = 644.752$, $p < .001$, $\omega^2 = 0.10$ |
| German | 3466 | 47.7 (39.3) | 674 | 36.6 (42.7) | There was a significant effect of license on monograph downloads, $F(1, 907.778) = 38.820$, $p < .001$, $\omega^2 = 0.01$ |
| Italian | 1258 | 37.1 (40.1) | 748 | 21.6 (24.9) | There was a significant effect of license on monograph downloads, $F(1, 2000.270) = 113.112$, $p < .001$, $\omega^2 = 0.04$ |
| Other languages | 97 | 63.5 (51.4) | 912 | 22.9 (22.1) | There was a significant effect of license on monograph downloads, $F(1, 99.828) = 59.341$, $p < .001$, $\omega^2 = 0.14$ |

Results: The assumption of homogeneity of variance has been violated; therefore, the Welch F-ratio is reported.

5.6.2.2 *Conclusions on the impact of license-enabled aggregation on downloads*

In contrast to the download activity prior to the launch of the Directory of Open Access Books, there is a statistically significant effect on all subsets: the use of an open licence, which allows the creation of an additional access point through the DOAB, has a positive effect on the number of books downloaded. The influence of aggregation clearly makes a difference. The most positive statistical effects are found within the subset “Sociology & anthropology” – where approximately 17% of the difference can be explained by open licensing and the subset “Politics & government” and the subset “Society & culture: general” – here approximately 12 % is measured.

However, not all results are so unambiguous, especially for the subsets on language. For instance, while a positive influence has been measured, the value of ω^2 for books in English is just 0.01. On the other hand, the mean

number of downloaded English language books on a *gratis* license is less than half the mean number of books on a *libre* license.

We can conclude that the use of *libre* licenses has a positive effect when we look at the effect of aggregation on downloads. Although the licenses do not directly affect the readers' behaviour, *libre* licences enable additional services by intermediaries like the Directory of Open Access Books. These additional services lead to increases in the number of books downloaded.

5.7 Discussion

The notion that *libre* material will be more used compared to *gratis* works seems highly obvious: an open license removes a barrier to usage. On the other hand, if the *gratis* works are made available under the same technical conditions as their *libre* counterparts, most users would make no distinction and treat the works as 'free as in beer'. In the case of the OAPEN Library, its description of licenses states the following: "If not stated otherwise, all works in the OAPEN Online Library fall under the OAPEN Deposit License – all rights reserved. End users are allowed to read the work online, download, print and copy it for their own personal purposes within the legal framework of their national copyright law. Beyond this all rights are reserved."¹⁵ In other words, the site clears legal obstacles for readers who want to use the books for personal reasons, and in this context it is not surprising that *libre* licenses did not play a large role in the period before the launch of the DOAB (January 2011-March 2012).

We have seen that each of the four discussed influences – *libre* versus *gratis* licenses; additional aggregation; subject and language – all affect the usage of the books in the OAPEN Library. By looking at the period before the extra coverage provided by the Directory of Open Access Books could play a role, a possible influence is removed from the analysis. As a second restriction, the usage data is split among subjects or languages. Within some of these subsets, the *libre* license positively affects usage, while in other subsets the effect could not be measured. However, even if a statistically significant result has been found, the effect size was negligible. The biggest measured impact of licenses found in the analysis of the subject subsets is approximately 2%. If languages are examined, almost 70% of all titles listed an effect of approximately zero percent. These results refute the claim by Guibault (2011) that open licenses enhance usage. However, in this

¹⁵ <http://oapen.org/about?page=support&subpage=forreaders>

particular case, the legal restrictions toward books with a more restrictive license are relatively slight.

Combining *libre* licenses and aggregation in the DOAB has a far more profound effect. When the data of that period is split in subsets based on subject or language the difference is clear. In each subset, the books with a *libre* license are downloaded more; the additional access provided through the DOAB appears to result in more successful dissemination of the books. This is also seen in the ratio between the mean number of downloads before and after the deployment of the DOAB. Taking into account all the average downloads in the subject subset reveals that in the pre-DOAB period, the number of downloads for books with a *gratis* license is 72% of the amount associated with books published under a *libre* license. After the launch of the DOAB, this percentage plummets to 43%. The same holds true in the language subset, where the percentages are 91% and 54%, respectively. This is another indication that extra aggregation has a positive impact on usage.

5.8 Conclusion

As far as could be established, this is the first paper to measure the effects of *libre* licenses on the use of open access monographs. Most of the literature on open licenses discuss them from a legal perspective, and focus on their innovations in relation to copyright. Also, open access publishing as a means to optimize the dissemination of scholarly and scientific information is mostly absent from the articles cited. However, the underlying theme – ownership and control over creative works and its economic aspects – does of course play an important role in the OA debate. Enforcing restrictions based on copyright laws creates another barrier to access, or to certain types of reuse.

Both the open access movement and the Creative Commons organization strive to maximise the use of creative works. While they share the goal of removing legal barriers to use or reuse, there is disagreement about the optimal license for open content. The Creative Commons organization chooses a flexible approach, by offering six different choices. In contrast, within the open access movement, there is a strong preference for the CC-BY license.

The current collection of the OAPEN Library does not completely conform to the recommendations of either group. Roughly half of the collection is made available under a *gratis* license that only permits personal use, which is more limited than the most restricted Creative Commons

license. Nevertheless, when considering direct use only (pre-DOAB launch downloads), the books under a *gratis* license perform just as well as the *libre* titles. In this context, the impact of licenses is limited.

However, when examining the use of OAPEN Library books after the launch of the DOAB, which automatically imports metadata of all books with a *libre* license, a benefit of *libre* licenses becomes clear. As Carroll (2006) predicted, machine readable metadata on licenses was used to perform a service; in this case inserting the OAPEN titles into the DOAB discovery service. Doing so proved to be successful: the titles featured in the DOAB are downloaded from the OAPEN Library more compared to books which do not receive the extra attention.

To a certain extent, the decision to include *libre*-licensed OAPEN titles in the Directory of Open Access Books – leading to additional visibility on another platform – has been a DOAB policy decision, and was not inherently dependent on license type. However, the machine readable *libre* licenses that enable aggregators such as the DOAB to identify and add licensed content can also lead to other types of reuse. For example, Biomed Central offers text mining services based on a collection of articles with a “BioMed Central open access license agreement”. According to BioMed Central, this license is identical to the Creative Commons Attribution License (BioMed Central Ltd., 2014).

Whether through simple aggregation or more intensive reuse like textual analysis, it appears that *libre* licenses do have the potential to positively affect usage. Rather than directly appealing to end users of individual books, these licenses enable intermediaries to create new services built on collections of open content. These services, in turn, can help to increase the impact of the individual publications.

5.9 Limitations

In the data set used for this paper, each book’s license was described in two ways: Creative Commons or no Creative Commons. It did not take into account the six different licenses in several versions – 2.0, 2.5 and 3.0 – that have been used in the examined collection. Some of the books were published under the UK or German version, while most were published under the ‘international’ version. It may be possible that the readers of the OAPEN Library were aware of all the legal details, and this influence has not been taken into account. The metadata of the books – available at <http://>

persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-8ut1-25 – contains the license of each individual title.

In the statistical analysis, it has been assumed that the choice for publishing a book under a *gratis* or a *libre* licence has not been biased. The influence of license on the behaviour of readers has of course been extensively discussed.

5.10 Acknowledgements

The author would like to thank professor Paul Wouters of the Centre for Science and Technology Studies (CWTS) for commenting on the draft version of this paper and Marieke Polhout of Data Archiving and Networked Services (DANS) for making the data available. Furthermore, the support of the publisher and the copy-editor have been crucial for the publication of this paper.