



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

## 2 Introduction

This dissertation will discuss the dissemination and usage of open access monographs, something that I have been working on since 2008. Here, open access monographs are defined as a scholarly piece of writing of book length on a specific subject, disseminated online in such a way that its contents can be read and downloaded without any barrier. Disseminating academic books in this manner is part of the open access movement, which aims to make scientific and scholarly content available to all. Peter Suber – considered to be the de facto leader of the open access movement – describes the rationale as such: “[R]esearch that is worth funding or facilitating is worth sharing with everyone who can make use of it.” (Suber, 2012).

Platforms for open access monographs are fairly new and they are just one aspect of the changes in the way scholarly and scientific results are made public. As I became involved in the development of both the OAPEN Library and the Directory of Open Access Books, questions on optimization arose. How can we improve these open access book platforms if there are few examples to learn from? An optimal solution should be based on evidence and my research on the dissemination and usage of freely available academic books aims to uncover relevant facts.

### 2.1 A short history of open access

Starting in 1991, preprints of physics papers were distributed using a central repository mailbox. The number of articles grew, and the repository expanded to include astronomy, mathematics, computer science, quantitative biology. In 2001, this repository was renamed to arXiv.org. The rise of the world wide web further enabled worldwide online distribution and in 2002 this idea was captured in the Budapest Open Access Initiative (BOAI) declaration (Chan *et al.*, 2002), where the term “Open Access” was coined. In the same year, the first set of Creative Commons licenses was released. These licenses enable the reuse of the contents in varying degrees. The role of licenses in the dissemination of open access books will be discussed in more detail in chapter 5.

At the start of the twenty-first century, several large scale open access initiatives were founded: PubMed Central<sup>1</sup> and the Public Library of Science.<sup>2</sup> Since then, other journal platforms such as PeerJ,<sup>3</sup> F1000Research<sup>4</sup> and Open Library of Humanities<sup>5</sup> have emerged. An important online book platform, the Google Books program, started in 2002.<sup>6</sup> A decade later saw the launch of several open access monographs platforms. In 2010, the OAPEN Library<sup>7</sup> was launched. In 2012, the Directory of Open Access Books<sup>8</sup> was introduced, listing monographs contained on several platforms. The next year, SciELO<sup>9</sup> and OpenEdition<sup>10</sup> started book platforms.

The introduction of new platforms for journal articles and books is part of a profound change in scholarly communication: the traditional roles of participants are changing. Some publishers are building their own digital collections, a task normally associated with libraries. On the other hand, academic libraries are starting up publishing activities (Bonn & Furlough, 2015), and publishers like Open Book Publishers or the Open Library of Humanities are led by academic authors. Lastly, some funders are managing their own collections, and – through crowdfunding – readers can finance books. For instance, the Austrian science fund FWF directly places books in the OAPEN Library (Snijder, 2015). Other funding bodies – such as the Spanish National Research Council – have chosen to set up an institutional repository (Bernal, 2013). The organisation Unglue.it uses a crowdfunding model to pay the rights holders of books to make them available through an open license. Among other types of books, academic books are part of the crowdfunding efforts (Howard, 2012).

## 2.2 Defining usage

Providing a general definition of “usage” is challenging; in this dissertation, the term “usage” as it refers to open access monographs is defined as

1 <https://www.ncbi.nlm.nih.gov/pmc/>

2 <https://www.plos.org/>

3 <https://peerj.com/>

4 <https://f1000research.com/>

5 <https://www.openlibhums.org/>

6 <https://www.google.com/intl/en/googlebooks/about/history.html>

7 <https://www.oapen.org>

8 <https://www.doabooks.org>

9 <http://books.scielo.org/>

10 <http://books.openedition.org/>

accessing the contents of the books. This is not exactly the same as reading a monograph. Most of this dissertation's research is done using the OAPEN platform. On that platform, it is not possible to measure whether a monograph has been read; instead the number of downloads is recorded. In a similar vein, the usage of the Google Book platform is measured as the number of pages that have been shown, or the number of times a book has been accessed. The results of the OAPEN Library and Google Books can be seen as a proxy for reading the books, but "flipping" a page in Google Books or downloading a book from the OAPEN library is no absolute guarantee that the person has actually read the monograph.

Many open access advocates stress the importance of reusing the contents of the scientific or scholarly documents that have been made available freely. This is supported by open licenses such as the Creative Commons licenses, which enable a certain amount of reuse by others. While the importance of reuse is not disputed, I will not discuss it in much detail. The primary reason is that reuse is even harder to measure than accessing content. At this very early stage in the development of open access monograph platforms, there are no reliable indicators available. This is not limited to reuse. For journal articles, measuring the number of citations is common practice. For monographs, this is not the case: chapter 9 describes the difficulties to obtain citations. Thus, in my definition of usage I have purposefully omitted reuse.

The question whether open access leads to more usage of monographs has already been settled in other research (Emery *et al.*, 2017; Ferwerda, Snijder, & Adema, 2013; Snijder, 2010). Making academic books freely accessible invariably increases the number of pages read online or the number of copies downloaded; a conclusion that is rather obvious. The next phase is to examine how to optimize that usage, and whether the increased usage has positive effects in academia and beyond.

The dissemination of open access monographs depends on platforms that offer a two-parts solution: a digital collection and the means of dissemination. When a platform is created, its administrators have to make decisions on what books to include. The collection as a whole will affect which users will be interested in using the platform, but we will also see that different aspects of the individual books affect the usage. Throughout the dissertation, the role of subject and language will be discussed in detail.

However, whether the platform reaches the intended audience depends not only on its contents. Just as important are the technical possibilities of the platform. Not just the question of how visitors can interact with the platform is significant, but also whether the contents can be integrated

into other environments. The impact of content integration will be made visible in chapter 4 and chapter 5.

### 2.3 Books versus journals

The difference in coverage between articles and monographs is visible in a recent review article on the impact of open access (Tennant *et al.*, 2016). It aims to list all current knowledge of this subject, but focuses only on journal articles as a way to publish scientific or scholarly results. However, monographs are an important publication type in the humanities and social sciences. Williams *et al.* (2009) conclude that “the monograph continues to enjoy unique appeal and status”, a clear indication of its standing.

Journal articles and monographs differ in several ways. The most obvious difference is the length: the average number of pages in an article is most likely around fifteen,<sup>11</sup> while the average monograph will contain around 300 pages. The latter publication form is clearly more suited for a thorough discussion of a subject. However, a longer text also changes the preferred format: while articles are mostly read digitally, there is still demand for paper books. In this light, it is understandable that publishers and librarians are interested in the combination of open access and paper versions. Chapter 3 describes my research into the influence of open access on the sales of paper copies.

The number of book titles and the number of journal articles differ wildly. This is illustrated by the Directory of Open Access Journals (DOAJ) and the Directory of Open Access Books (DOAB). In August 2017, the DOAJ lists over 2.5 million articles. In contrast, the DOAB contains close to 8,900 titles. This difference has economic consequences. Articles tend to be more standardized, and due to concentration of publishers, economies of scale can be more easily achieved. In contrast, monographs tend to be treated like unique projects, and are published by a much larger number of publishers, considerably differing in size.

The difference in text length also leads to a different pace of interaction: it takes longer to write a monograph than it takes to create an article. Using citation analysis based on what is common in journal articles will not lead to optimal results. Any citation analysis on academic books, such as the research in chapter 9, has to accommodate for this. If the long “citation cycles” are problematic, other forms of assessment might be examined: for

11 See for instance Falagas, *et al.* (2013); Stremersch, Verniers, & Verhoef (2007)

instance, by looking at the usage data of open access monographs. This idea is further investigated in chapter 7.

## 2.4 Central thesis and research questions

In the introduction, I described my involvement with the OAPEN Library and the Directory of Open Access Books. Ultimately, these platforms aim to share the contents of freely accessible books as widely as possible, which is measured by the level of usage. While the usage of open access monographs depends on the removal of paywalls, the level of usage is primarily determined by other factors. Properties of the books such as language and scholarly field determine the possible readers and the way dissemination platforms are configured affect whether those readers can actually be reached.

The question which factors affect the use of open academic books is quite open-ended. In this dissertation, I will examine three main aspects: economic sustainability, optimisation of the infrastructure and evaluation of the results. Economic sustainability of open access monograph publishing is one of the basic conditions for the platforms: without books, there is no need for a platform. This leads to the question whether open access has a positive influence on the sale of monographs. For decades, the uneasy financial situation surrounding publishing academic books has been known as “the monograph crisis”. Decreasing sales and rising costs are threatening the economic sustainability of monograph publishing and publishers are exploring alternative business models. One of these is the so-called “hybrid model”, where an online version is made freely available, and paper copies must be purchased. Will the improved visibility lead to more sales? This is explored in chapter 3.

In addition to the economic aspects, I have examined the factors affecting the dissemination of open access monographs. Understanding these factors helps to optimize the platforms. A fundamental question for the development of both the OAPEN Library and DOAB is how to present the collection to prospective readers. Should the platform only be accessible as a “silo”, or should it try to integrate its offering in other systems? The answer to this question has consequences for the design. The “silo” approach assumes that humans reach the platform and start searching there, while system integration requires standardized book metadata that can be imported into the systems of libraries and aggregators. Chapter 4 deals with this question.

The optimization of open access platforms is not just dependent on technical choices. Another thing to consider is collection choices. In the case of the OAPEN Library, the collection contains books with a license that permits reuse and books with a more restrictive license. Does this difference in licensing affect the use? I have compared the usage of the two sets of books. Within the open access community, licenses that enable reuse of scholarly content are seen as very important, and chapter 5 examines whether this preference is also shared by the users of the OAPEN Library. Furthermore, the influence of aggregation through another platform – the Directory of Open Access Books – is measured.

Apart from licenses, the users of the OAPEN Library and DOAB may have other preferences. Understanding those preferences is useful to improve the platforms, but users are not required to register. Thus, no information about individuals is stored. The question is how to emulate the successful tactics of online retailers – that store the preferences of their clients – without violating privacy. A possible solution can be found in deploying social analysis techniques to discover user communities. See chapter 6.

The next chapters discuss the results of open access monographs dissemination; starting with the question of how to evaluate the effects of open access monographs. I have examined the possibility to quantify the effects of Humanities and Social Sciences (HSS) research, in a way that is relatively effortless. The results of chapter 7 aim to achieve this goal through the investigation of usage data.

If the goal of open access is to make research available to everybody, does it help to overcome the digital divide between the “global north” and the “global south”? Using open access book platforms requires a functioning digital infrastructure, which might set back readers in developing countries. Does open access lead to more usage in developing countries? The answer can be found in chapter 8. Lastly, I have examined if there is an “open access advantage” for monographs. It is widely documented for journal articles, but the effect of open access on citations is largely unknown. The same holds true for social media. Chapter 9 discusses the question whether open access monographs are cited more and receive more attention on social media.

I have conducted multiple studies on the usage of open access monographs, which are presented in the following chapters. Each chapter reviews a different aspect: book sales, digital dissemination, open licenses, user communities, measuring usage, developing countries and the effects on citations and social media.