



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

## 4 Modes of access : The influence of dissemination channels on the use of open access monographs

Snijder, R. (2014). Modes of access: the influence of dissemination channels on the use of open access monographs. *Information Research*, 19(3), 166–183.

Retrieved from <http://www.informationr.net/ir/19-3/paper638.html>

### 4.1 Introduction

Open access is much debated and in recent years has gained much attention in the literature. The scientific and scholarly impact of papers has been discussed extensively, for instance by Antelman (2004), who finds that freely published papers receive more citations – across a number of disciplines. Podlubny (2005) takes the citation analysis a step further and proposes a normalisation procedure, aimed at comparing the impact of scientists from different fields. Bollen *et al.* go beyond citations and investigate 39 impact measures, and conclude that “usage-based measures” may be a better indication of scientific impact (Bollen, Van de Sompel, Hagberg, & Chute, 2009).

Not only is the impact hotly debated but the economic aspects have also received much attention. A major discussion point is the merits of publishing a free version of a paper next to the ‘official’ version in a journal which is not freely accessible (green open access), versus the merits of directly publishing in an open access journal (gold open access) (Harnad *et al.*, 2004, 2008). Recently, the report *Accessibility, Sustainability, Excellence: how to expand access to research publications* by Finch *et al.* was heavily discussed (Finch *et al.*, 2013).

The discussion on the effects of open access on monographs does not attract the same amount of attention so far, and the amount of available research is small. Apart from running the OAPEN Library, the OAPEN foundation is currently involved in two pilot projects in the Netherlands (“OAPEN.nl website - English,” n.d.) and the UK (“JISC - OAPEN-UK,” n.d.) experimenting with open access monograph publishing. The first results of the OAPEN-UK pilot are discussed by Collins & Milloy (2012). In September

2013, the results of the Dutch pilot project were published (Ferwerda *et al.*, 2013).

## 4.2 Dissemination channels

This paper will focus on a different aspect: dissemination channels. In the literature on open access, dissemination channels seem to be a given. If it is discussed at all, dissemination is described as making papers available in an institutional repository. This paper is the first to analyse the effects of several dissemination channels in an open access environment.

Here we examine the monograph downloads of the OAPEN Library, which was officially launched in September 2010 (OAPEN Consortium, 2011; Open Access Publishing in European Networks, 2010b). It is a Web based collection of monographs, mainly in the field of Humanities and Social Sciences (HSS). All books are available in open access and users can search the Website in several ways. Each book also has a unique Web address and can be downloaded directly without searching the Website. These addresses – combined with metadata describing the books – are made available on the OAPEN Website and through several aggregators. This is described in more detail in Snijder (2013a).

This paper examines the download data of the OAPEN Library, which was gathered during a period of six months. The data consists of the number of downloads per month per provider. Here we define a provider as the organisation that grants the user access to the internet. Furthermore, the data contains information on whether a book was downloaded via the OAPEN Website or directly. Because the data were aggregated monthly, we can distinguish three situations: firstly, a book was downloaded a certain number of times through a provider via the Website only; secondly, a book was downloaded a certain number of times through a provider using the direct download address of the book; thirdly, a book was downloaded a certain number of times through a provider via the Website and also a certain number of times directly. In the last case, the readers related to that provider use a combination of ways to access the book.

It is not unreasonable to assume that each provider caters for several people. In the case where all readers only use the Website or only use direct downloads, their preference seems to be aligned. If – in the same month – a portion of the readers use the Website and another portion of the readers prefer direct downloads, this may hint at another 'group configuration'. In this case, other aspects of usage could also differ, which is why this is

analysed separately. Thus, the download data stems from three channels: Website only, Website and direct access; and direct access only.

As the data is available through several channels, it may be useful to investigate the literature on multichannel management. This field looks at the challenges that retailers face in the deployment of multiple channels to reach their customers. While typical research in this field looks at the differences between offline channels such as stores and online channels such as Websites, parts of the theoretical framework could be applied to this paper.

The multichannel management framework is based on theories on the adoption of innovations, explaining if and why people will use new channels. On this layer, the specific aspects of working with multiple (retail) channels are discussed. According to Rogers (1995), several factors influence the use of innovations: the relative advantage of the innovation, its fit with existing usage patterns, the perceived complexity, the ability to try out the innovation, the perceived risk related to adoption, and the degree to which adoption and use can be observed by others (Rogers, 1995).

The work of Rogers is paired to the technology adoption model (TAM) and its extension TAM2. TAM states that perceived usefulness and perceived ease of use are drivers of innovation adoption; TAM2 extends this framework to social influence processes (subjective norm, voluntariness, and image) and cognitive instrumental processes (job relevance, output quality, result demonstrability, and perceived ease of use) (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989; Venkatesh & Davis, 2000). Neslin *et al.* identified five “key challenges” in multichannel management: data integration, understanding customer behaviour, channel evaluation, allocating resources across channels and coordinating channel strategies. In a later paper, the list of relevant aspects has grown to thirteen (Neslin *et al.*, 2006; Neslin & Shankar, 2009). Basically, the questions revolve around whether or not to deploy a multichannel strategy, how to set up different channels, and how to evaluate the results.

What aspects of multichannel management can be used here? Instead of offline versus online channels, we are discussing different online channels. We envision different users with different needs. They are not paying customers, and researching and ‘purchasing’ in an open access environment are more or less the same action. Searching for information in the field of humanities and social sciences is covered by many authors. Shen discusses the many channels used by social scientists, grouping them in internal and external electronic and paper resources, combined with “external human resources” Shen (2007, p.8). Bulger *et al.* discuss humanities

scholar's search behaviour through six use cases where scholars employed a range of resources and technologies (Bulger *et al.*, 2011). Wang *et al.* use an international angle by discussing the scholars in the USA, Greece and China (Wang, Dervos, Zhang, & Wu, 2007). Griffiths and Brophy focus on students' online search behaviour, and describe the strong preference for search engines – especially Google – compared to the library catalogue or other sources (Griffiths & Brophy, 2005). Lamothe discusses the growing usage of e-books in an academic library (Lamothe, 2010).

Channel evaluation also has implications for resource management: the results help to decide where to invest the most time and money. This goes beyond managing IT systems, it also affects marketing decisions. In short, multichannel management aims to create an optimal strategy in a given environment.

If we combine search behaviour with the decision to use a specific channel, we arrive at the following research question: does the usage based on the channel “Website only” differ from usage based on “direct access only” or from usage from a combination of those channels? The answer has implications for open access publishing as it may help to optimise the dissemination of open access monographs.

First, the download data is analysed quantitatively: counting the number of downloads per channel. Then, the qualitative analysis tries to find an answer to the question of whether properties of the users, their infrastructure or the properties of the book themselves have a significant impact on the usage per channel.

### 4.3 Quantitative analysis

In this section, the data set is described, followed by the number of downloads per channel. The number of monograph downloads is an indication of readership. Whilst we can assume that the more a monograph has been downloaded, the more it has been read we cannot, however, state that 100 downloads equal equates to 100 people reading the monograph cover to cover.

#### 4.3.1 The data set

The data set consists of the download data of 979 books, published by thirty-five different publishers. The books are published in ten different languages. By far the largest number of the downloaded books are in English. The 979

monographs in the data set were downloaded 152,662 times in the first six months of 2012.

**Table 1 Languages: number of titles**

Language	Number of titles	Percentage
English	514	52.5%
German	164	16.8%
Dutch	125	12.8%
Other languages	176	18.0%
Total	979	100%

The ratios of the downloads per language are more or less in line with the percentages of published languages. This is discussed in more detail in the qualitative analysis. Annex 2 contains the complete list of languages.

**Table 2 Languages: number of downloads**

Language	Downloads	Download percentage
English	8,8003	57.6%
German	3,2632	21.4%
Dutch	1,9025	12.5%
Other languages	1,3002	8.5%
Total	152,662	100%

The following table lists the ten most downloaded subjects. This is a fraction of all available subjects: the complete data set contains eighty-three different subjects. The classification used is the BIC standard subject categories (BIC) (Book Industry Communication, 2010). The question of whether language or subject has a measurable influence on channel usage will be discussed in the qualitative analysis.

**Table 3 Subjects: most downloaded**

Subject	Number of titles	Percentage
History (HB)	165	17.0%
Politics & government (JP)	148	15.3%
Society & culture: general (JF)	80	8.2%
Sociology & anthropology (JH)	62	6.4%
Film, TV & radio (AP)	32	3.3%
Literature: history & criticism (DS)	37	3.8%

Subject	Number of titles	Percentage
Philosophy (HP)	25	2.6%
Religion & beliefs (HR)	23	2.4%
Science: general issues (PD)	34	3.5%
Laws of Specific jurisdictions (LN)	32	3.3%
Other subjects	332	34.2%
Total	979	100%

As before, the ratios of the downloads per subject are more or less in line with the percentages of published subjects. This is discussed in more detail in the qualitative analysis. Annex 3 contains the complete list of subjects.

Subject	Number of downloads	Download percentage
History (HB)	23,624	15.5%
Politics & government (JP)	19,167	12.6%
Society & culture: general (JF)	13,520	8.9%
Sociology & anthropology (JH)	9,033	5.9%
Film, TV & radio (AP)	6,571	4.3%
Literature: history & criticism (DS)	6,786	4.4%
Philosophy (HP)	5,896	3.9%
Religion & beliefs (HR)	4,506	3.0%
Science: general issues (PD)	3,796	2.5%
Laws of Specific jurisdictions (LN)	7,002	4.6%
Other subjects	52,761	34.6%
Total	152,662	100%

We saw that the 979 books were downloaded 152,662 times in the first six months of 2012. The books were accessed through 6176 different providers which are based in 166 countries. We stated before that a provider is defined as the organisation that grants the user access to the internet. In some cases, the provider is an organisation such as a university or a government agency. In other cases, this is an Internet Service Provider (ISP), such as Comcast in the USA or Ziggo in the Netherlands. The providers will be discussed in more detail in the qualitative analysis.



### 4.3.2 Downloads per dissemination channel

The downloads were measured per provider per channel per month. So, if a provider downloaded the same monograph more than once in the same month, using the same channel, the number of downloads were added. In some instances, a provider downloaded a monograph several times a month via the Website and also via direct access. In those cases, the downloads were added to the combined channel “Website and direct access”. In other instances, a monograph was only downloaded via the Website, or the monograph was only downloaded through direct access only. Then the downloads were added to the channels “Website only” or “direct access only” respectively.

Using this procedure, the following data becomes available:

Channel	Number of downloads	Percentage
Website only	11,546	8%
Website and direct access	29,453	19%
Direct access only	111,663	73%
Total	152,662	100%

The data shows that usage is dominated by direct access only. This implies that almost three quarters of all downloads come from users who do not use the Website <http://www.oapen.org/>, but find the books via other means. This kind of usage is made possible by making the metadata<sup>12</sup> of the books – including a direct download URL – directly available to all interested parties, including libraries and content aggregators. The channel “Website and direct access” contains a combination of downloads via the Website and direct access. Here again, the portion of downloads via direct access is larger than the downloads via the Website. It is clear that most readers find the books through other routes than the OAPEN Library Website.

The usage data revealed that 24% of the visits to the OAPEN Library Website lead to downloading one or more titles. However, this percentage cannot be compared to the usage data of other systems. If 100 OAPEN monographs were downloaded via a library catalogue, how many searches were conducted which did not result in a download taking place? Therefore, we do not know whether the OAPEN Library Website is a more efficient way to search compared to other systems.

<sup>12</sup> The metadata is licensed under a Creative Commons Zero licence, which makes it free to use under any circumstance.

We discussed before that multichannel management aims to create an optimal strategy in a given environment. The goal of open access publishing is to remove barriers to access, and it makes sense to investigate how to maximize the dissemination of open access monographs. We saw that the direct access channel is far more used than the other channels and this has serious consequences for managing and optimising the service: from a dissemination point of view it makes more sense to invest in metadata and the dissemination of metadata then to spend resources on the Website. It is important that any system used for open access dissemination is capable of exporting metadata in formats that can be used by content aggregators or the systems used by prospective readers. Apart from library catalogues, search engines may be a much-used research tool, and investing resources in optimal coverage by the likes of Google and Bing may be beneficial.

#### 4.4 Qualitative analysis

The goal of the qualitative analysis is to establish whether user's characteristics (i.e. their infrastructure) or the collection are influential factors on channel usage. Firstly, user characteristics are discussed. The download percentages of the quantitative analysis are used as a benchmark, and are compared to the actual values found using an independent t-test. A factor is considered influential if the difference between the usage numbers is statistically significant and the effect size is not small.

##### 4.4.1 Characteristics of users and dissemination channels

Readers are placed in several groups: academic; government; business; non-profit organisations and the general public. While academic users could be seen as the main audience for monographs, readers of other backgrounds have equal access to the monographs in the OAPEN Library. The users are categorised based on the data from the OAPEN logs, combined with public data.

The OAPEN Library is a Web based service, and its logs contain the Web address of the providers. So, if researchers at Leiden University download a book using their office equipment, the Web address ([www.leidenuniv.nl](http://www.leidenuniv.nl)) of that university will be logged. Basic information such as address and telephone number are publicly available and can be found using the so called 'WHOIS protocol' ("WHOIS - Wikipedia," n.d.). By combining the usage data and information about the provider, we can make assumptions about who is downloading a specific monograph.

A large portion of the providers are not universities or government agencies, but internet service providers (ISP). If the provider is an ISP, the user cannot be linked to an organisation. We cannot assume that all usage through an ISP comes from people browsing the internet at home. If the internet infrastructure in a certain country is highly developed, chances are that each organisation is capable of giving direct internet access to their members. If the internet infrastructure is less well developed, a large portion of the organisations in that country do not directly provide internet access but rely on the services of an Internet Service Provider.

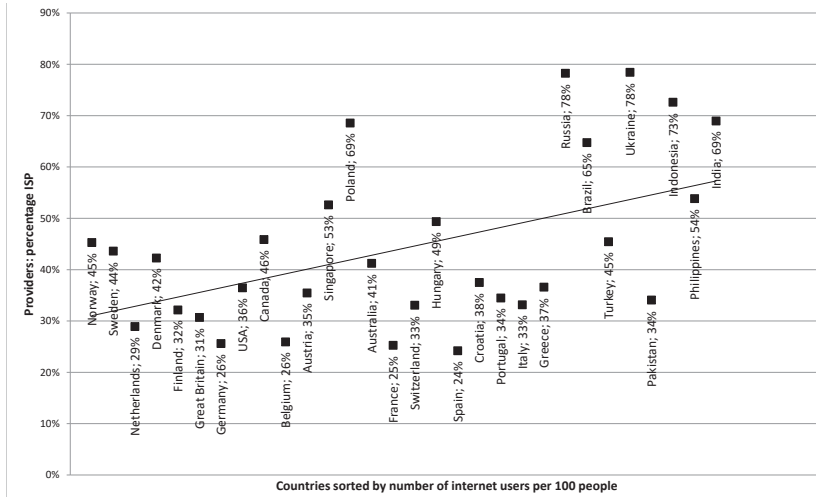
Of course, it is always possible that 'ISP users' from a country with a highly-developed internet infrastructure are in fact academics working from home after office hours. The available data does not contain the (local) time of the download, which makes determining whether a reader is downloading during office hours impossible. Furthermore, if the reader is not acting in a professional capacity, the chances are also higher that the download started after office hours. The difference in access to scholarly and scientific literature for academics compared to others is quite large; using the credentials of the academic institution allows direct access to all kinds of literature behind pay walls. It might therefore be more efficient to use these credentials not only at the office, but also after office hours.

If we want to divide ISP usage in those two categories, we need a way to determine the state of a country's infrastructure. This is done by using a World Bank publication: *The Little Data Book on Information and Communication Technology 2011* (World Bank & Lewandowski, 2010). It lists several statistics per country, and one of them is the number of internet users per 100 people. If there is a connection between the state of the infrastructure and the percentage of downloads through ISPs, the percentage of 'ISP usage' is lower for highly developed internet infrastructures.

This assumption was tested by charting the measured downloads from thirty countries and the percentage of ISP usage. The found values were set against the amount of internet users per 100 people. Because the country of each provider is known, it was possible to select the countries with the highest number of downloads. The selected thirty countries are responsible for almost 92% of all downloads.

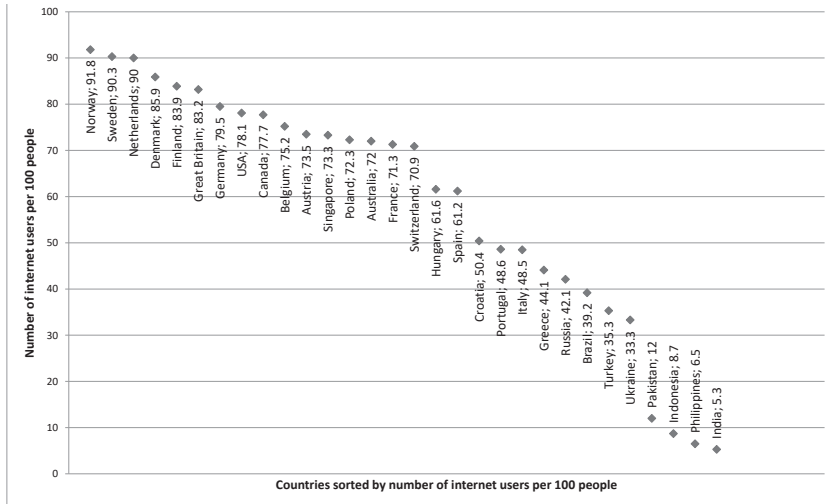
The first chart depicts the percentage of downloads via an ISP, sorted by the amount of internet users per 100 people. In this chart, we see that there is a trend toward a higher percentage of downloads via an ISP, when the number of internet users per 100 people decreases.

Figure 1 ISP percentage



The second chart depicts the number of internet users per 100 people. Here we see a decrease from 91.8 internet users per 100 people in Norway to 5.3 internet users per 100 people in India. Somewhere between these two extremes we need to set a cut-off point to determine which countries have a highly-developed internet infrastructure. Within these countries, the chances are higher that ISP usage from these countries is from “non-professional” users. This distinction is used in the qualitative analysis, to determine whether the internet infrastructure influences downloads via the different channels.

**Figure 2 Internet users per 100 people**



between Switzerland with 70.9 internet users and Hungary with 61.6 internet users per 100 people. Based on this, the threshold is set to seventy internet users per 100 people. Countries with seventy or more internet users per 100 people are considered to have a highly-developed infrastructure. The same threshold is also used in (Snijder, 2013b).

#### 4.4.2 Type of users and dissemination channels

Now we can look at the download percentages of the different user groups. The number of downloads per channel differ wildly and because of that, there is a large difference in the absolute number of downloads per group. For instance, the number of downloads by academic readers through the “direct access only” channel is almost seven times the number of academic downloads through the Website only.

Is there a connection between user type and dissemination channel? Regardless of the channel, most of the usage comes from three groups: Academic, ISP and ISP high internet usage. As academics are the intended audience for monographs, it is not very surprising to see a large proportion of usage that originates from academic institutions. Furthermore, the academic community is large. As discussed before, it was not possible to determine whether the role of users in the group “ISP” was academic or otherwise. The members of the group “ISP – high internet” are more likely to be ‘non-professional’ users. Based on that we might conclude that disseminating open access books helps to make scholarly content available

to the public. In all channels, the usage by non-profit, government and business organisations is small, compared to academic and ISP related usage.

From the quantitative analysis it becomes clear that 8% of the usage comes from the channel Website only, 19% from the channel Website and direct access, and 73% through direct access only. We can use these percentages as a baseline for the expected downloads for each user group, and compare it to actual number of downloads per channel. Using the difference between those amounts – expressed as the percentage of the expected value – we find no significant effect for user type:  $t(17) = -0.541$ ,  $p = .595$ . Based on the lack of significant differences on channel usage, we can conclude that the type of user plays a minimal role in channel usage.

Type of user	Website only		Website and direct access		Direct access only		All channels Actual
	Actual	Expected	Actual	Expected	Actual	Expected	
Academic	3,005	2,312	5,821	5,490	20,068	21,092	28,894
Business	49	393	11	933	4,849	3,583	4,909
Government	162	171	17	406	1,959	1,561	2,138
Non-profit	136	121	31	287	1,346	1,105	1,513
ISP - high internet usage	4,138	6,134	14,000	14,567	58,531	55,968	76,669
ISP	4,056	3,082	9,573	7,323	24,910	28,134	38,539
Total	11,546	12,213	29,453	29,006	111,663	111,443	152,662

#### 4.4.3 Characteristics of internet infrastructure

Dividing the internet structure in highly developed and less well developed countries is not only useful to differentiate between user groups but is, in itself, also a possible influence on channel usage. We might expect that readers from countries with a highly-developed infrastructure have different download patterns compared to those with more limited bandwidth. Annex 1 lists the countries with highly developed infrastructure. When we look at overall usage – not taking into account the different channels – the difference between the two groups is clear: the number of downloads from countries with a highly-developed infrastructure is more than twice the number of downloads from the rest of the world.

The same percentages as before are used as a baseline for the expected downloads, and again those numbers are compared to the actual number of downloads per channel. Using the difference between those amounts – expressed as the percentage of the expected value – we find no significant effect for internet infrastructure:  $t(5) = -0.418$ ,  $p = .639$ . Based on the lack

of significant differences on channel usage, we can conclude that internet infrastructure plays a minimal role in channel usage.

Internet infrastructure	Website only		Website and direct access		Direct access only		All channels
	Actual	Expected	Actual	Expected	Actual	Expected	
Less than high usage	5,051	3,768	12,226	8,948	29,817	34,378	47,094
High usage	6,495	8,445	17,227	20,058	81,846	77,065	105,568
Total	11,546	12,213	29,453	29,006	111,663	111,443	152,662

#### 4.4.4 Characteristics of content and dissemination channels

Is there a connection between characteristics of the content – the monographs – and dissemination channels? In this section, we examine two aspects: subject and language. Not all languages or subjects will be analysed: the three most downloaded languages and ten most downloaded subjects are examined.

#### 4.4.5 Language and dissemination channels

It seems obvious that language influences the use of the monographs, as readers are unlikely to download a book in a language they cannot read. The high usage of monographs in the English language is directly visible, but we have to take into account the large number of books available in that language. The question is whether language usage differs significantly from expected values.

In the description of the data set, we saw that 52.6% of the books were written in English, 16.7% in German and 12.9% in Dutch. If we apply these percentages to the number of downloads per dissemination channel, we can compute the expected values. Using the difference between those amounts – expressed as the percentage of the expected value – we find no significant effect for language:  $t(11) = -1.229$ ,  $p = .245$ . Based on the lack of significant differences on channel usage, we can conclude that language of the monographs does not play a role in channel usage.

Language	Website only		Website and direct access		Direct access only		All channels
	Actual	Expected	Actual	Expected	Actual	Expected	Actual
English	9,808	6,073	20,389	15,492	57,806	58,735	88,003
German	471	1,928	4,472	4,919	29,318	18,648	32,632
Dutch	396	1,489	2,843	3,799	14,157	14,405	19,025
Other languages	871	2,055	1,749	5,243	10,382	19,876	13,002
Total	11,546	11,546	29,453	29,453	111,663	111,663	152,662

Still, the percentage of downloads of English language books via the Website is relative high, and this raises the question of whether users primarily search using English terms. To test this, a small sample was analysed. Of all queries in one month, a list was created of searches that occurred at least twice. This created a set of 2,219 different queries.

	Number of queries	Percentage
In English	1,074	48.4%
Not in English	1,145	51.6%
Total	2,219	100%

The percentage of 'non-English' queries was more than 51%. Nevertheless, this group also contained search terms that exist not only in the English language, but also in Dutch and German. If we analyse this group, five 'ambiguous' terms account for more than 62% of queries: "film"; "water"; "IMISCOE"; "Iran"; "Islam". So, a large percentage of all the examined queries are at least 'compatible' to English. It is therefore safe to assume that most searches are indeed in English, which would partly explain the results. The large amount of available English language books might be another factor.

Multilingual terms	Number of queries	Percentage
"film"	348	30.4%
"water"	167	14.6%
"IMISCOE"	150	13.1%
"Iran"	31	2.7%
"Islam"	24	2.1%
Other terms	425	37.1%
Total	1,145	100%



#### 4.4.6 Subject and dissemination channels

The last aspect to analyse is the subject of the monographs. Are the users of the OAPEN Library interested in certain subjects or do the download patterns closely follow the spread of subjects amongst the books? We have found the percentages of titles with a certain subject in the quantitative analysis. The expected number of downloads per channel are computed by applying these percentages to the number of books downloaded per channel, and the actual number of downloads are compared against the benchmark values. Using the difference between those amounts – expressed as the percentage of the expected value – we find no significant effect for subject:  $t(32) = 1.507, p = .142$ . Based on the lack of significant differences on channel usage, we can conclude that subject does not play a role in channel usage.

Subject	Website only		Website and direct access		Direct access only		All channels <i>Actual</i>
	<i>Actual</i>	<i>Expected</i>	<i>Actual</i>	<i>Expected</i>	<i>Actual</i>	<i>Expected</i>	
History (HB)	1,751	1,963	5,647	5,007	16,226	18,983	23,624
Politics & government (JP)	1,649	1,767	3,713	4,506	13,805	17,084	19,167
Society & culture: general (JF)	1,512	947	2,645	2,415	9,363	9,156	13,520
Sociology & anthropology (JH)	1,061	739	1,900	1,885	6,072	7,146	9,033
Film TV & radio (AP)	869	381	1,422	972	4,280	3,685	6,571
Literature: history & criticism (DS)	324	439	1,340	1,119	5,122	4,243	6,786
Philosophy (HP)	304	300	1,413	766	4,179	2,903	5,896
Religion & beliefs (HR)	367	277	879	707	3,260	2,680	4,506
Science: general issues (PD)	505	404	605	1,031	2,686	3,908	3,796
Laws of Specific jurisdictions (LN)	134	381	485	972	6,383	3,685	7,002
Other subjects	3,070	3,949	9,404	10,073	40,287	38,189	52,761
Total	11,546	11,546	29,453	29,453	111,663	111,663	152,662

However, when a dissemination channel is used more – the channel “direct access only” is used for of 73.1% of all downloads, while usage through the Website only is 7.6% – the number of subjects also grows. This is illustrated by the fact that the ten subjects listed here cover almost 74% of

all downloads occurring via the Website only. In contrast, the percentages drop for the other channels to 63.8%.

## 4.5 Conclusions

This paper is the first to analyse the effects of several dissemination channels in an open access environment. Its goal is to help determine an optimal strategy to achieve maximum distribution of open access monographs. The books are made available via the OAPEN Library Website, via direct downloads or a combination of those two. It is interesting to note that a large proportion of readers who directly download the monographs do not use the Website; they have found the description of the books via other means.

From the quantitative analysis, the dominance of one channel is clear. The data shows that 73% of all downloads occurred via the channel “direct download”. This implies that almost three quarters of downloads come from users who do not use the Website <http://www.oapen.org/>, but find the books through other systems or Websites.

The qualitative analysis revealed that regardless of the channel, most usage comes from three groups: Academic, ISP and ISP high internet usage. Other user groups – Business, Government and Non-profit – are not highly represented. When looking at the usage per group, no effect on channel usage could be established. The internet infrastructure is another factor that was taken into account. While the digital divide between users from countries with a highly-developed internet infrastructure and users from less well-off countries is very clear, no effect on channel usage could be found. The same holds true for the aspects of the books themselves: the analysis could not find any effect on channel usage for both the language and subjects of the monographs.

The goal of multichannel analysis is to determine the optimal usage of resources: what configuration leads to the best results? The definition of “best results” in an open access environment differs from a commercial environment. The objective is not financial gain, but maximum dissemination. In the OAPEN Library, readers can access books via three channels. Firstly, the Website, which is optimised for search – it does not only contain metadata, but also enables full text search. Furthermore, it contains browsing functions as a means to enable serendipity. In contrast to this, the direct search channel functions in a different way. It is based on metadata only, which is incorporated into systems outside the OAPEN Library. Full text

search on the contents of the books is not possible. The third channel is a combination of both.

The results show that most readers are using the “direct download” channel, in spite of the fact that the OAPEN Library Website offers functions that are not available via other channels. A possible answer may be found in the theoretical models on the use of innovations discussed in the introduction. There we saw several factors influencing the use of new systems, such as its fit with existing usage patterns, perceived ease of use and social norms. It is possible that users of the “direct download” channel prefer their ‘own’ systems, which are familiar and are part of their routine and environment. In that case, learning to use a new interface may not be seen as a worthwhile investment. But who are the principal users of the OAPEN Library? The analysis revealed that current users are based in academic institutions or use an ISP. Users based in businesses, governmental or non-profit organisations are far less common. Also, the digital divide between upcoming countries and the ‘developed’ countries is a large factor: two-thirds of the downloads occurred from countries with a highly-developed internet infrastructure. And although the OAPEN Library contains books in German, Dutch, Italian and other languages, the majority of the books – and the majority of the readers – use English.

How does this compare to the goal of maximum dissemination? A recurring theme in the discussion on open access is making scientific and scholarly results available to members of academia who cannot access the information behind a pay wall. Seen from that perspective, the current situation is quite a success: academic institutions are responsible for a large portion of the downloads. However, when we look at other possible patrons, the picture is less rosy. In the collection of the OAPEN Library, the subjects politics & government, society & culture and sociology & anthropology are well covered. Those books may contain useful information for governmental organisations – for instance in the field of immigration studies, which is a much-debated topic in Europe and North America. Nevertheless, there is not much usage from governmental organisations, nor from non-profit organisations. Does the form – monographs – not fit within the informational habits of those potential users, or is the OAPEN collection not embedded in the information systems used?

When we compare the usage from countries with a highly-developed internet infrastructure to the usage from the rest of the world, the difference is striking. The first group of countries contains twenty-seven countries, yet it has downloaded twice as much as books. Here we see that making books freely available does not automatically take away other barriers to access.

The language of the publications may be another issue to research. More than half of the analysed books are written in English, and the download percentage of English language books is also roughly 50%. It is possible that the overall usage is at least partly shaped by the number of books available in a certain language. In other words, if the collection contained a larger percentage of monographs in another language – for instance French, Spanish and Portuguese – how might that affect the usage?

The results imply that making the metadata available in the user's systems – the infrastructure used on a daily basis – ensures the best results. So, to achieve the optimum amount of usage, firstly we must identify users who are not using the data; secondly, we have to understand how they search for information and thirdly we have to establish what is the best way to make our data available. Researching those questions would bring the goal of maximum dissemination a little closer. These challenges are not only faced by the OAPEN Foundation, but are shared by all organisations that disseminate open access publications or data.

#### 4.6 Limitations

The data set used in this paper is large and rich: it contains the data of 979 monographs which were downloaded 152,662 times in the first six months of 2012. Several aspects of the monographs are described: language and subject. Furthermore, several characteristics of each download are available: the name of the provider and the channel used.

Nevertheless, as there is no authority data obtainable, the categorisation of the providers is not checked. Another aspect linked to the categorisation of providers is determining its country of origin, based on the available WHOIS information which always links one country to a provider. If an organisation operates in several countries – such as a NGO or a multinational – this will not be reflected in the data. Also, the subject information of the books has been simplified. These aspects may have had an influence on the qualitative analysis.

The timeframe could also be considered. The data was captured during a six-month period, and owing to the rapid pace of technological development on the internet, it would be interesting to compare the results with data from another period. Because this research is the first of its kind, no best practises have been established.

## 4.7 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.

## 4.8 Annex 1: list of countries with a highly-developed internet infrastructure

According to *The Little Data Book on Information and Communication Technology 2011*, the following countries have 70 or more internet users per 100 people:

- Andorra
- Australia
- Austria
- Belgium
- Bermuda
- Brunei
- Canada
- Denmark
- Estonia
- Finland
- France
- Germany
- Great Britain
- Iceland
- Japan
- Luxembourg
- Netherlands
- New Zealand
- Norway
- Poland
- Singapore
- Slovakia
- South Korea
- Sweden
- Switzerland
- United Arab Emirates
- USA

#### 4.9 Annex 2: downloads per language

Language	Downloads	Percentage
English	88,003	57.6%
German	32,632	21.4%
Dutch	19,025	12.5%
Italian	8,586	5.6%
Danish	1,387	0.9%
French	629	0.4%
English, Latin	594	0.4%
German, Latin	488	0.3%
Spanish	476	0.3%
French, Latin	395	0.3%
German; English	236	0.2%
Norwegian	115	0.1%
Welsh	96	0.1%
	152,662	100%

#### 4.10 Annex 3: downloads per subject

Subject	Downloads	Percentage
History (HB)	23,624	15.5%
Politics & government (JP)	19,167	12.6%
Society & culture: general (JF)	13,520	8.9%
Sociology & anthropology (JH)	9,033	5.9%
Film, TV & radio (AP)	6,571	4.3%
Literature: history & criticism (DS)	6,786	4.4%
Philosophy (HP)	5,896	3.9%
Religion & beliefs (HR)	4,506	3.0%
Science: general issues (PD)	3,796	2.5%
Laws of Specific jurisdictions (LN)	7,002	4.6%
History of art / art & design styles (AC)	4,092	2.7%
Humanities (H)	3,810	2.5%
Society & social sciences (J)	3,439	2.3%
linguistics (CF)	3,317	2.2%
Economics (KC)	3,118	2.0%
Literature & literary studies (D)	2,415	1.6%
Industry & industrial studies (KN)	2,252	1.5%
Business & management (KJ)	2,117	1.4%
The environment (RN)	1,848	1.2%
Law (L)	1,673	1.1%
Biology, life sciences (PS)	1,566	1.0%
Theatre studies (AN)	1,352	0.9%
Library & information sciences (GL)	1,350	0.9%
Interdisciplinary studies (GT)	1,348	0.9%
Architecture (AM)	1,334	0.9%
Archaeology (HD)	1,283	0.8%
Psychology (JM)	1,034	0.7%
Economics, finance, business & management (K)	956	0.6%
International law (LB)	907	0.6%
Education (JN)	891	0.6%
The arts: general issues (AB)	851	0.6%
Digital lifestyle (UD)	820	0.5%
Industrial chemistry & manufacturing technologies (TD)	650	0.4%
Music (AV)	600	0.4%
Educational material (YQ)	578	0.4%
Jurisprudence & general issues (LA)	531	0.3%
... (HJ)	487	0.3%
Poetry (DC)	391	0.3%
Social services & welfare, criminology (JK)	371	0.2%
Language (C)	361	0.2%
... (JR)	361	0.2%
Medicine (M)	358	0.2%

Subject	Downloads	Percentage
ELT background & reference material (EB)	328	0.2%
Warfare & defence (JW)	296	0.2%
Romance (FR)	287	0.2%
Agriculture & farming (TV)	285	0.2%
Prose: non-fiction (DN)	282	0.2%
Earth sciences (RB)	266	0.2%
Memoirs (BM)	248	0.2%
Biography & True Stories (B)	228	0.1%
Sports & outdoor recreation (WS)	227	0.1%
Civil engineering, surveying & building (TN)	224	0.1%
Finance & accounting (KF)	218	0.1%
... (HF)	204	0.1%
Adventure (FJ)	203	0.1%
Fiction & related items (F)	188	0.1%
... (QM)	185	0.1%
Graphical & digital media applications (UG)	182	0.1%
Encyclopaedias & reference works (GB)	181	0.1%
Medicine: general issues (MB)	172	0.1%
Databases (UN)	168	0.1%
Geography (RG)	161	0.1%
Mathematics (PB)	158	0.1%
Language teaching & learning (other than ELT) (CJ)	155	0.1%
Crime & mystery (FF)	144	0.1%
... (DV)	127	0.1%
Antiques & collectables (WC)	115	0.1%
Local interest, family history & nostalgia (WQ)	113	0.1%
Earth sciences, geography, environment, planning (R)	97	0.1%
Reference, information & interdisciplinary subjects (G)	89	0.1%
Art treatments & subjects (AG)	87	0.1%
Environmental science, engineering & technology (TQ)	87	0.1%
Modern & contemporary fiction (post c 1945) (FA)	79	0.1%
Other branches of medicine (MM)	77	0.1%
... (JB)	75	0.0%
Astronomy, space & time (PG)	68	0.0%
... (LK)	66	0.0%
Museology & heritage studies (GM)	59	0.0%
Biography: general (BG)	56	0.0%
Computer science (UY)	44	0.0%
Fiction: special features (FY)	39	0.0%
Art forms (AF)	24	0.0%
... (JS)	8	0.0%
	152,662	100%