



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

10 Conclusions

10.1 Introduction

The deliverance of open access monographs is a complex process. It is based on the actions of many different stakeholders, who have invested time, money and other resources in order to make academic books freely available online. The monographs are disseminated through several platforms, which are part of a larger online ecosystem containing search engines, library catalogues, social media platforms and many more components. Potentially, everyone who is connected to the internet can access the books. The new technical possibilities enabled publishers and funders to offer online collections, while empowering librarians and authors to publish books.

In this dissertation, the usage of open access monographs is seen as an indication of success. However, the actions of the stakeholders, the complexities of the online ecosystem and properties of the monographs could all affect the usage. To simplify the discussion, I will first discuss the properties of academic books that are not related to online dissemination. In this way, it is possible to make a distinction between aspects that are tied to the concept of the book – whether published in digital or paper form – and the aspects that are connected to the virtual realm. For instance, language is an aspect of the book that affects the usage on online platforms: books in other languages than English are mostly used by native speakers.

I assume that the content of the monograph is created independently of its appearance: the paper version of the book contains the same information as the online version.¹ This assumption has an important consequence: the changes brought on by open access consist largely of adjustments in the online infrastructure. The development of new platforms such as institutional repositories, Google Books, the OAPEN Library etc. are examples of such changes. The performance of these platforms depends partly on their technical specifications, partly on the books themselves.

In the majority of cases, up until the moment the book is ready for publication, the workflow is still firmly grounded in the traditions of the print era; see for example Springer's workflow (Springer, 2017). For some

¹ There are exceptions to this rule. An example is the book "Vincent van Gogh – The Letters" and the accompanying website <http://vangoghletters.org>.

authors, the new possibilities offered by digital publishing could be used to critically engage with the current publication model. Alonso *et al.* (2003) discuss abandoning the print model in favour of the digital possibilities. Hall (2013) is going even further by questioning the concept of the book itself. Instead of a bound entity, the 'new' book is in constant flux: updated through the engagement of researchers and others. At this moment, most books – digital or not – are still a far cry from this vision: a stable text-based publication, consisting of chapters and pages (Carmody, 2011). For now, the current online publication form is basically a digitized version of the paper copy; the same holds true for most journal articles, which also did not change in a significant way (Ware & Mabe, 2015).

Throughout this dissertation, books are considered to be stable objects, which are not inherently changed by open access dissemination. Several aspects of the book, however, will affect online dissemination. Two aspects have been examined in the previous chapters: language and subject. A third aspect is more implicit: quality in connection to trust. In this concluding chapter I will look in more detail at quality and trust, after a short discussion of language and subject.

The influence of language on dissemination is profound. An author who wants to reach a global audience needs to publish in English. Any other language than English will mainly attract a "local" audience, such as Germany, Austria and Switzerland in the case of German language books, or the Netherlands and Belgium when Dutch language books are on offer. This can be inferred from chapters 6 and 7. This bias towards English also extends to citation indexes, a topic that will be discussed in section 10.5.

In this dissertation, the subject matter of a book is used as a proxy for scholarly discipline. The scholarly discipline's influence can be found in two areas: dissemination and assessment. In short, subject defines the audience: most of those who are interested in film and media studies are not trying to acquire expertise in the field of archaeology. Bibliometric methods such as citation counts can be seen as a form of assessment. For the humanities and social sciences, this is not without problems (Nederhof, 2006; Ochsner *et al.*, 2017). For instance, each scholarly discipline has different citation practices; which is visible in section 10.5.2. In addition, the results of chapter 9 seem to suggest differences in Twitter mentions per scholarly discipline.

Scholarly research is diverse. Ochsner *et al.* (2017) provide a useful summary of common characteristics. Research in the humanities and social sciences might attempt to accumulate knowledge in the same linear fashion as the natural sciences, or it might be focused on interpreting and reflecting on existing phenomena, such as texts and theories in the humanities and

concepts in the social sciences. In other words, instead of striving towards one definitive answer, it tries to create new perspectives and thus works with competing visions. The researched phenomena can be local, such as the history of a specific region. This also leads to the use of 'local' languages, instead of English.

The diversity in research practices is also reflected in citation culture and quality assessment: each scholarly discipline has different norms whether a publication has sufficient quality. However, consensus exists about one aspect: a publication's quality should not be solely determined by the author. Sufficient quality can only be determined by the author's peers. Both the Royal Netherlands Academy of Arts and Sciences (2010) and Ochsner *et al.* (2012) conclude that quality is best assessed by researchers in the same field, assisted by additional indicators. One of those indicators could be the prestige of the publisher of monographs, as discussed by Giménez-Toledo & Román-Román (2009).

The notion of research quality is determined to a large extent by the scholars within a discipline. Additionally, funders and publishers play a part as well. Funding agencies can influence the research agenda by deciding which research – or scholar – receives subsidy. Moreover, when the role of funding agencies also encompasses publishing open access content – for instance by demanding an open publication license or by using an open access platform – they directly shape the publishing landscape. Publishers play a similar role by deciding whether to accept a manuscript, and by enabling dissemination through open access channels.

Related to the quality of the publications is the issue of trust. Most readers and the libraries and aggregators that act on their behalf will validate the online books on offer. Do they have confidence in the book and the platform it resides on? For instance, when an author publishes a monograph on a personal website, will it reach the same level of usage compared to the same book published on the publisher's website? Intermediaries such as academic libraries might place more trust in the publisher's offering (Moghaddam & Moballeghi, 2007). Another aspect of specialised platforms is their optimisation towards online usage. Not just search engine optimisation, but also by offering services to the intended audiences.

In conclusion, quality assessments directly affect the dissemination of open access books, through the combined actions of the stakeholders. Some groups may act as gatekeepers, strongly affecting the diffusion of books. This is illustrated in chapter 5, where listing titles in the Directory of Open Access Books (DOAB) enhances usage. Before the launch of DOAB, all titles in the OAPEN Library saw comparable levels of usage. When a

set of titles were listed in DOAB – purely based on the licence and not on the contents or the quality of the books – their usage soared compared to the unlisted titles. It seems more than likely that intermediaries accepted DOAB as a valid source, resulting in additional exposure. For instance, the growth of DOAB is listed in the “Dramatic growth of open access” blog (Morisson, 2016). This is further illustrated by McCollough (2017), who sees the Directory of Open Access Books as a tool for discovering open access monographs in academic libraries.

In a sense, the dissemination of books is the final phase of the publication cycle. However, publishing monographs is financially challenging and in section 10.3 I discuss my research on the economic sustainability of a specific model: hybrid publishing. Open access books are disseminated using several platforms, and section 10.4 discusses the optimisation of the infrastructure. After the books have been disseminated, the question arises how to evaluate the results. My answers can be found in section 10.5.

10.2 Web based data sets and data providers

In this section, I will briefly discuss some properties of the analysed data sets. With the exception of the data set of chapter 3, the data have been selected using the web. Collecting data in a web environment is almost by definition automated, eliminating manual procedures and enabling the creation of large data sets in a relative effortless way. However, it also poses challenges. As the environment changes constantly, the gathered data is strongly connected to a certain period in time. For example, the estimated number of websites in 2010 was 200 million, in 2017 the number grew to over 1.7 billion (“Total number of Websites - Internet Live Stats,” n.d.). This is also true for the OAPEN Library itself: the number of titles grew from slightly over 850 titles in 2011 to 2,300 books in 2014. In July 2018, the collection comprises almost 5,500 titles. Not just the number of titles increased, also the number of users and the number of book downloads, leading to possible changes in interaction: changes in user’s countries; changes in providers and aggregators; changes in the collection’s subjects and languages.

On top of this, online tools change or disappear. The data gathered for chapter 8 is based on geographical data provided by Google Books; since 2012, this platform has stopped offering this type of data. The research on monograph citations of chapter 9 is based on Google Scholar. In the year after that research was completed, Google Scholar decided to index the contents of the OAPEN Library (Pinter, 2015). Whether this affects the

number of citations found, is not known. The same chapter also used the services of the Topsy search engine to find tweets. The Topsy.com service has been discontinued in December 2015. Thus, replicating the research on a later date is hardly possible. This is a known problem that affects all researchers working with web based data.

The data sets can be divided in three groups. The data of chapter 3 consists of the sales data of all books published under the same imprint by Amsterdam University Press. The data of chapter 4 to 7 is based on the logged usage data of the OAPEN Library, combined with the metadata describing the books and – where applicable – the added metadata describing the providers. Here, the selection of books is based on all books that were part of the collection during the period under scrutiny. In contrast, the data sets of chapter 8 and 9 are based on a curated and much smaller selection of books. Here, the experimental set and the control set are chosen carefully to remove bias. The data to be analysed is derived from web based platforms: Google Books, Google Scholar and Topsy.com.

When the size of the data sets is compared to the sets used in bibliometric research – for instance in Costas *et al.* (2014); Thelwall *et al.* (2013) – the number of titles is small. Also, the data has mostly been derived from one platform: the OAPEN Library. This might lead to a certain amount of bias. Yet, even the smallest data set is based on nearly 200 books, which are selected carefully to *remove* bias. The larger data sets are based on hundreds of titles, published by dozens of publishers, spanning multiple subjects and several languages. On top of that, the influence of language and subject is analysed separately from the possible effects of open access. Comparing usage data from other platforms would be a good way to enhance our understanding, but comes with its own challenges: differences in infrastructure, collection or definitions of usage must all be accounted for.

I have applied several analytical techniques to the different datasets. In numerous occasions, I applied analysis of variance (ANOVA) to determine whether the influence of one of more aspects of the books or the book's users is more than coincidental. In order to produce reliable outcomes, the values in the data set must be distributed normally; the so-called "bell curve". When the values are out of kilter compared to a normal distribution – which is the case in chapter 9 – I have used the generalised linear model (GLM). The most recent research – described in chapter 6 – was based on social network analysis, combined with a clustering algorithm. The conclusions derived from these analyses will be discussed in more detail in the next sections.

10.3 Economic sustainability

Books are the result of a network of organisations and individuals working together. This network has to be economically sustainable. However, the economic sustainability of monographs has been problematic for decades, long before the advent of open access book publishing. The introduction of chapter 3 describes falling print runs, declining sales and shrinking budgets in academic libraries.

When financing books in a commercial setting is far from easy, how are the costs met if the books are made available for free? What business model can be applied? Publishing monographs in open access could be seen as a “system break” (Pochoda, 2013) or a transition from print-only to digital – mostly in combination with printed books. According to Adema & Ferwerda (2014), this opens new possibilities: increased dissemination, combined with new possibilities to search the contents of collections of books. However, in order to reach this state of affairs it is necessary to find a business model that works.

Several business models for academic books have been discussed in the literature. Greco & Wharton (2008) recommend looking into a model optimized for open access books, combined with a print on demand system, for those who still prefer a paper version of the book. The search for new business models is also described by Withey *et al.* (2011), who are investigating how to preserve the best elements of the current publishing system in a new era of open access monographs. There are numerous other business models, ranging from a hybrid publication model to crowd-funding (Ferwerda, 2014). Recently, Knöchelmann (2017) discussed the open access book market, tying successful upscaling to funding.

Within the direct sphere of influence of the OAPEN Library and the Directory of Open Access Books (DOAB), different business models are used. For instance, the French organisation OpenEdition – which makes titles available via DOAB – offers a “freemium” package to libraries: a combination of a basic version of a publication that is freely available online, combined with paid-for premium services (Mounier, 2011). Knowledge Unlatched is another example, using a model based on the cooperation with university libraries. It establishes a library consortium that pays a “Title Fee” to a group of publishers. In return, the publishers offer print copies to member libraries at a discount and also make the books available in open access. One of the deployed platforms is the OAPEN Library (Pinter, 2012).

Very few papers can be found on the costs of producing monographs. The costs of creating a – paper only – monograph by an American university

press is discussed by Wasserman (1998). Roughly speaking, the costs of around \$24,000 are not met by the expected sales: depending on the edition, the losses range between \$8,000 and \$13,000. A more recent investigation by Ferwerda *et al.* (2013) looked into the publication costs of Dutch monographs publishing. Based on the budgets of 50 books the average costs for publishing a monograph in the Netherlands was found to be slightly over €12,000. And finally, Maron *et al.* (2016) examined the publication costs of 20 American publishers in 2014. The average costs of a digital monograph ranged from \$30,000 to \$49,000. Whether each amount is based on the same cost structure is unknown. Recently, Pinter (2018) explains that comparing monograph publications costs is problematic, due to the diversity of publishers.

It is doubtful whether the economic sustainability of monographs is guaranteed by the sales of paper copies, and the literature on costs seems to suggest that a substantial amount of money is needed to produce an academic book. In such circumstances, will open access publishing have a positive monetary effect?

Chapter 2 offered some further insight into the economic effects of open access monograph publishing, by examining the effects of a hybrid business model. In such a model, paper copies of books are sold, while an online version is also made available for free. The main assumption is that the open access version of the title acts as an “advertisement”: when the reader has discovered the book online, this will possibly lead to the purchase of the paper version, as many readers still have a strong preference for the paper codex. As a counterargument, one might argue that paper books are not a necessity in the era of e-book readers and high-quality tablets. The main question is thus whether the hybrid business model enhances or diminishes sales.

The data underlying chapter 3 does not come from a controlled environment, such as described in Snijder (2010). Instead, it examines sales data obtained from a “normal” business setting: sales data from Amsterdam University Press obtained in the period 2010 to 2012. While the publisher uses the hybrid business model – selling paper copies alongside online open access versions – the commercial expectations for the open access titles differ from the closed access titles. This can be inferred from the print run: a higher print run indicates a higher expectation of number of copies sold. The average print run of books published in closed access was much higher, compared to the open access titles. Apart from commercial potential, the moment of sales is also an important factor: most copies are sold in the first year of publication.

Chapter 3 set out to measure the influence of open access on monograph sales. Furthermore, the effect of open access was compared to other influences on monograph sales: commercial potential, front list and back list, and language. Each influence is statistically relevant, making it harder to single out the effects of open access. The difference in number of copies sold in the first year – the front list – compared to the number of copies sold in the subsequent years – the back list – is striking: the mean sales in the first year is about five times larger than the year after that. Consequently, I analysed the front list sales and the back-list sales separately.

The results of the front list sales can be explained by a combination of commercial potential and language; open access publishing does not have an effect in this situation. The results for the back list are similar to the front list outcomes. The influence of language was not statistically relevant, and open access publishing is a relevant influence on sales in certain cases only: the subsets of books whose print run is between 1 and 2000. The resulting average number of copies sold seems to point to a small advantage for the closed access titles. Whether the advantage of closed access published books is economically relevant, is questionable. Over 65% of all copies sold were open access titles.

In the debate on the economic sustainability, the small differences in the number of copies sold are not the main issue. In all discussed experiments open access did not have a large effect on monograph sales, positive nor negative. At the start of this section, I mentioned the problems in the book trade, and I have found that the hybrid model does not lead to more sales.

10.4 Factors affecting dissemination

So far, I have discussed the aspects of the books which remain stable in a paper and a digital environment and the financial fundament under monograph publishing. The next aspect to explore is online dissemination. The distribution of open access monographs consists of two parts: a digital collection and the means of dissemination. In the previous chapters, several platforms were introduced: institutional repositories, publisher's collections, the Google Books platform, the OAPEN Library and the Directory of Open Access Books (DOAB). Some aspects of this non-exhaustive list will be summarized in this section, as an illustration of the open access monographs infrastructure.

Each platform has its own affordances. For instance, disseminating books via an institutional repository may underline the relation with the hosting

organisation. The Google Books platform enables different things: besides being directly linked to the Google search engine, it allows rights owners precise control over how much of the book is made visible to the public. A platform such as the OAPEN Library is optimized for disseminating OA books via several channels. The Directory of Open Access Books only stores metadata, but amplifies the use of the titles listed.

Understanding the strengths and weaknesses of platforms is vital for choosing a dissemination strategy. Online dissemination platforms shape what the readers can do with the book, which affects its usage. The usage data generated by the platform can be used to assess the impact of the books on the platform, an idea investigated by Herb *et al.* (2010) and in the previous chapters. What a platform is capable of, is decided by its owner. Each owner will have different preferences, leading to a landscape of various possibilities. To illuminate the differences, I will shortly discuss the platforms and their owners.

Institutional repositories are based on a set of standards promoting interoperability. Each repository should be able to connect to other repositories and use its content. They could be seen as a natural extension of academic libraries: in most cases the library will manage the repository. Other platforms are also used within the library community: some librarians make a part of their collection searchable through Google Books (“Library Partners – Google Books,” n.d.). Platforms such as the OAPEN Library or DOAB are also used as a source for OA books. Apart from academic libraries, some funding institutions may choose to directly deploy repositories or comparable platforms. For instance, the Austrian science fund FWF directly places books in the OAPEN Library (Snijder, 2015). Others, for instance the Spanish National Research Council, have chosen to set up an institutional repository (Bernal, 2013).

Some publishers – for example Brill or ANU Press – have made a collection of books available on their website. Setting up a bespoke platform enables publishers to control what data to collect about the users. Some people will argue that knowing more about the people active on a platform solely benefits the platform owner. A recent example is the speculation by Kelyt (2016) about the motives of Elsevier to purchase the SSRN platform: SSRN’s data can be used as a means to evaluate scholars; to be sold to university administrators. The question of privacy is discussed further in chapter 6: how to balance the privacy of the readers versus the desire to know the “customer” in detail? A publisher might also use other platforms to distribute open access books online: Google Books, OAPEN Library or DOAB.

Strictly speaking, the Google Books platform is not an open access platform. It is a search engine that contains and indexes books, which also allows the rights owner to decide how much of the book's contents is publicly visible. This feature enables publishers to fully open up a book if desired. Controlling the visibility of the book's content can be used to set up experiments in which a set of books with a limited amount of visible content is compared to a collection of books where all pages were visible (Snijder, 2010). However, publishers and libraries do not control the platform, and the platform's owners decisions may not always suit them. For instance, since 2015 no new publishers are allowed to sign up to the Google Book platform (S. Hall, 2016).

The platforms differ in capabilities, but also in content. Each platform strives to maximise its use – at least within its target audience – and a major factor is the quality of the offering. Thus, I assume that each platform will select suitable titles and refuse inappropriate ones. What is a suitable collection will be different for each platform: institutional repositories and publisher's platforms will be limited to their organisations; the OAPEN Library and DOAB collect titles from different publishers but emphasize quality assurance of the titles; the Google Books platform attempts to keep pirated books from their collection.

Maintaining a trusted platform might also be a strategic advantage for the hosting organisations. For publishers, it may be a way to directly sell copies to readers – cutting out the middle man. For academic libraries, it may be a way to strengthen their position within the university, and a possible counterweight to the influence of publishers. In the case of large commercial organisations, the platform may be part of other offerings. The success of Google depends at least partly on knowing the preferences of their users. The kind of information gathered may lead to privacy concerns. This conflict of interests has been discussed in more detail in chapter 6.

Open access book platforms are still a relative new phenomenon. What aspects are important for the dissemination of open access books? In the next section, I will discuss several of these aspects.

10.4.1 What works in digital dissemination?

The research in the previous chapters is based on experiments, carried out on several platforms. Most experiments have taken place using the OAPEN Library. The OAPEN Library has been operational since 2010, making it one of the longest running open access monograph platforms. It has several properties that help examine the influences on the usage of

monographs. Firstly, its collection of several thousand books contains large groups of books in several languages, especially English, German, Dutch and Italian. Furthermore, the collection spans a broad range of subjects. The monographs are not only available through the OAPEN interface, but – through availability of metadata and agreements with commercial and non-commercial aggregators – are also directly accessible via library catalogues and other platforms. Due to the fact that the platform has been operational for several years, trends over longer periods can be examined. The diversity in licences is another factor that can be studied. Lastly, the books made available on this platform have been vetted through a peer review process.

Before, I have discussed economic sustainability as a basic requirement for disseminating open access books. Now I will look into another aspect affecting the distribution of open access monographs: dissemination channels. Online dissemination contains more than placing documents on a website, hoping they will magically turn up prominently in the results of search engines. Instead, it is necessary to use the channels that are best suited for the targeted audience. Until recently, 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.

In chapter 4, the success rate of two dissemination modes has been examined: the OAPEN website acting as an Online Public Access Catalogue (OPAC), and direct access where the reader directly downloads the book without searching the website. A “direct” download implies that the reader has used other means to find and select the book. The direct search channel is based on metadata only, which is incorporated into systems outside the OAPEN Library. The usage data obtained comes from three channels: through the website only; a combination of website and direct downloads; or downloads only. The data is analysed both quantitatively and qualitatively in chapter 4. The quantitative analysis reveals a large difference between the number of books that were downloaded without searching the OAPEN website and the other dissemination channels: 73 % of all downloads can be attributed to ‘direct’ downloads. The results of the qualitative analysis are not so easy to interpret: the provider’s characteristics nor the properties of the books were statistically significant.

The books were downloaded through providers, which I categorised in two ways: the type of provider and the state of their country’s internet infrastructure. This categorisation was introduced in chapter 7, which will be discussed later. The question is whether a connection between

the provider category and dissemination channel exists. Regardless of the channel, most of the usage comes from three types of provider: academic, internet service providers (ISP's) and ISP's from a country with a highly developed internet infrastructure.

The state of development of a country's internet infrastructure does not affect which dissemination channel is used. However, the digital divide is clearly visible in the smaller usage from the countries with a less-developed internet infrastructure compared to the small group of better equipped countries. Lastly, the subject or language of the downloaded books did not affect the usage of the channels.

A possible explanation for the large percentage of direct downloads can be found in the theoretical models on the use of innovations. Whether or not a new system is used depends on several aspects, such as its fit with existing usage patterns, perceived ease of use and social norms. It is possible that most users prefer their 'own' systems, with which they are familiar and which are part of their routine and environment. In that case, learning to use a new interface may not be seen as a worthwhile investment.

The high percentage of direct downloads – over 70% of all book downloads – cannot be fully explained by search engine optimisation, as only 30% of the internet traffic to the OAPEN Library during that period originated from search engines. This means that a sizable portion originated from other types of websites. The only way to directly download the books is by using a specific download address. Those addresses are distributed by the OAPEN Library, through its metadata feeds. When other systems or websites incorporate the web addresses that enable direct downloads of books, they act as aggregators. While I did not examine this, it is likely that some websites only display a portion of the collection. An example is the Ancient World Online blog (Jones, n.d.), which lists only monographs about the Antiquities Period.

Before, I stated that the success of open access publishing depends on many stakeholders. The main purpose of open access is to make knowledge available, and it is useful to investigate the factors that enhance dissemination. The results of chapter 4 reveal an important aspect of open access dissemination: enabling incorporation into other systems enhances the monograph usage. Here, the solution offered by the OAPEN Library is providing metadata to be used by aggregators. While the metadata is available to all, a relative small portion of usage can be attributed to search engines. The indexation by search engines is an automated process, but the incorporation of the metadata into other systems – which aggregate information for readers – is the result of a conscious decision. I conclude

that this decision is based on trust. Aggregators accept the monographs offered by the OAPEN Library as a viable source, and make them available to their patrons.

The importance of aggregators is also visible in the results of chapter 5: their influence on usage is much stronger than that of licenses. Within the literature on open access, the role of licenses is discussed extensively. According to the Open Access Scholarly Publishers Association (OASPA), “true” open access can only be achieved through the use of a specific Creative Commons license: CC-BY (Redhead, 2012). If true open access means optimal dissemination of scholarly content, books published under an open license – which allows sharing its contents – should perform better than books made available under a license that permits nothing more than downloading for personal use. I tested this hypothesis on the OAPEN Library, where roughly half of the collection is available under a license permitting reuse, and the other half under a license that only permits personal use. The results showed that the number of downloads of open licensed books did not differ significantly from the monographs with a “free to read” license.

However, I also investigated the role of the Directory of Open Access Books (DOAB), by examining the usage data of the same collection after the launch of DOAB. The DOAB aggregates open access books, but only those with an open license. Open licenses such as Creative Commons are machine-readable: they can be used in automated processes, leading to new possibilities. In the case of the OAPEN Library, the licensing information is part of the metadata. The metadata is used by the DOAB, in order to select books with an open license. When the period after the launch of DOAB was examined, the difference is far greater. Books listed in the DOAB have been downloaded almost twice as much on average compared to the other group of titles. Even when allowing for the role of subject and language, the influence of DOAB is profound.

While the license is seen by many in the scholarly communication field as an important enabler for open access, it is doubtful whether the readers care as much. The results seem to suggest that a “free to share” license is not an important incentive compared to a “free to read” license. The number of downloads was not boosted by an open license, the usage was boosted by incorporation of a new service: DOAB. It is DOAB policy to only list monographs with an open license, and thus half of the OAPEN Library collection was imported, leading to the large difference in usage.

The influence of other aggregators could explain the large uptake of the books listed in DOAB. When more aggregators are aware of the existence of DOAB, compared to the OAPEN Library, the monographs listed in DOAB

will receive more attention. Several authors see DOAB as a comprehensive source of open access monographs (McCollough, 2017; Morisson, 2016). More exposure will also lead to more data usage. Again, usage is strongly affected by trust: being listed in DOAB – a widely trusted source – results in more aggregation and thus more visibility, which stimulates the usage of open access monographs.

As far as usage by readers is concerned, the results of chapter 5 seem to downplay the role of licenses. Given the fact that a large percentage of the books were published under a CC-BY-NC-ND license – which does not permit commercial use or creation of derivative works – other stakeholders might consider those licenses as equal to ‘free to read’ licenses. For instance, research institutes may be more strongly bound to the terms of the licenses, especially when a large set of books is examined. The use of large corpora for text mining depends on permissions by rights holders (Van Noorden, 2014). Still, the influence of aggregation in DOAB is undeniable: even the books published under the most restricted open licence have been used more, compared to the books available for reading purposes only.

In conclusion, while licenses are important for certain groups of users, this is not the case for those who want to read the books. For them, usage is not boosted by licenses, but by the choices of aggregators.

10.4.2 Clustering books and readers

So far, I have looked at book dissemination purely based on numbers; examining factors affecting the number of downloads, a proxy for the number of times a book is read. Chapter 6 uses a different angle: creating clusters of books that are suitable for a group of readers. Instead of lumping the users of the OAPEN Library together into large groups such as academics, government employees or the general public, an attempt is made to uncover “communities”: groups of people that share an interest. Defining communities and finding suitable titles is an important task of libraries. Online retailers such as Amazon use a different strategy, based on personal recommendations. Creating a more fine-grained understanding of the users of any open access platform helps to deliver the best titles. However, it also leads to questions of privacy: is it desirable to store information about individuals? These questions are examined in chapter 6.

One of the most prominent success factors of online retailers is the amount of knowledge they possess about their customers. If the preferences of each client are known, it is possible to offer desirable products. In such circumstances, the online retailer will strive to maximise the amount

of known facts about all their customers. I noted before that collecting and storing data about individuals leads to discussions about privacy. For libraries, the protection of their patron's privacy is an important part of their core values. Online dissemination platforms could model themselves after online retailers; after all, apart from charging money for their services, they perform more or less the same functions. However, the main purpose of open access platforms is not to maximise sales, but to maximise the usage of documents, which is closer to the core values of libraries.

I investigated whether it is possible to create optimized recommendations while storing a minimum amount of information about individuals. A solution for this problem might be found in the download behaviour of all users of a dissemination platform. By analysing all data at once, instead of focusing on individuals, it might be possible to discern patterns: clusters of related books that are downloaded together. If such clusters can be found, they could form the basis of a recommendation, akin to recommendations by online retailers. To create an optimal solution, it is also necessary to understand who is interested in a specific cluster of books, without targeting individuals. To resolve this, the research focuses on finding communities: groups that share a common trait.

The research was based on two data sets, consisting of providers, books and the number of times a book was downloaded. The first set was captured during 2012 and the next set is based on data from 2014. Each book in the collection was categorised through its language and subject. The information about providers is limited to name and country of origin. The linked titles and providers are clustered using the Wakita-Tsurumi (2007) algorithm, resulting in dozens of clusters. The ten biggest clusters were analysed, comparing the books' language and subject and the providers' nationalities to the complete data set.

Within the examined data, several clusters could be identified that were not the result of random downloads. Some clusters contain large percentages of non-English books, combined with a large set of providers consisting of native speakers. An example is a cluster containing Dutch language books combined with many providers from the Netherlands and Belgium. Other clusters – where the language is mostly English – contain books on certain subject, such as film and media studies or Indonesia and South-East Asia. When the subject is region-based, this is also reflected in the nationality of the providers.

The clusters are not created manually, but are the result of an algorithm. Consequently, this procedure can be part of an automated process, akin to the recommendation services of online retailers but without violating the

privacy of individuals. However, there is still room for improvement: clusters found in 2012 are not visible in the 2014 data. This is not uncommon: other research on clustering techniques also show differences, all of which might be valid in their own right (Gläser *et al.*, 2017).

In conclusion, to a certain extent it is possible to use clustering algorithms to create optimized recommendations, while still protecting the privacy of individual readers. Optimized recommendations by open access platforms should lead to higher usage of open access monographs. The results of chapter 6 can be seen as a proof of concept, to be further refined.

10.5 Evaluation of results

Until now, I discussed the hybrid business model and several aspects of digital dissemination affecting the usage of open access monographs. From these practical considerations I will now move to the outcomes: does publishing monographs in open access lead to a greater scholarly impact and societal influence? To answer this question, I first need to define scholarly and societal impact, insofar as it applies to monographs. Open access monographs can have an impact on the work of academics – I will categorize this as academic or scholarly impact – and they might affect those who do not have access to large academic libraries – defined here as social or societal impact.

Monographs require other indicators than journal articles. Bibliometric measurements like the journal impact factor have been used for decades (Garfield, 2006). For monographs, similar data is not abundantly available; instead, metrics based on library holdings might be used. My research is based on usage data, derived from online platforms. In this case, the proxy value for academic impact is the amount of usage originating from academic institutions, compared to usage from other organisations. This metric is restricted to the number of academics who use the internet infrastructure of their institution to access the OAPEN Library; it will not take into account academics who use other internet providers. My research on the academic impact of open access monographs is not limited to usage data: in chapter 9 I have examined whether open access affects the number of citations.

I have examined the social impact of open access monographs using indicators based on usage data. When the usage originates from governmental, non-profit or business organisations, I have classified this as types of social impact. Another indicator of the social impact of monographs can be found in altmetrics, here defined as online activity about academic publications.

Some types of online activity are closely tied to the work of academics, for instance Mendeley.com or ResearchGate.com. Others – such as Facebook or Twitter – are used by a large section of the general public. Mentions of open access monographs on those platforms stand a larger chance to come from non-academics.

In short, in this section three types of indicators will be discussed: citation based indicators, platform usage and altmetrics. The definition and mutual relations of these indicators is discussed in more detail by Glänzel & Gorraiz (2015), who state that the combination of usage, altmetrics and citations leads to a more complete view of a document's impact. According to the authors, citations are an accepted indicator of academic impact, but do not capture social impact. Usage measures the intention to read documents and altmetrics indicate mentions of documents, both in academia and beyond.

10.5.1 Impact measured

Indicators of academic impact are relatively easy to identify through usage originating from academic institutions. Social impact is more diverse: it encompasses usage by non-academic readers with a professional interest such as government employees, but also readers without a professional interest: members of the general public. To distinguish between these groups, I use the connection to an organisation – which can be inferred from the usage data – other than an Internet Service Provider or an academic institution. I assume that non-academic readers with a professional interest are connected to an organisation.

The defining characteristic of members of the general public is their lack of connection to an organisation. This complicates identification based on usage data: if readers use an Internet Service Provider (ISP), does that mean they are not connected to an organisation, does it mean that “their” organisation is unable to provide direct internet access, or are they just not using their organisation's equipment? Differences in internet infrastructure are also at the root of the digital divide between developing and developed countries, leading to the question whether open access leads to more usage when the available internet infrastructure is not optimal.

Categorizing users in groups is useful to distinguish between the usage by academics and usage by others. Simply put: usage of monographs by non-academics is a form of social impact. Comparing the percentage of non-academic users of a set of open access monographs to a set of monographs in closed access helps to determine whether open access leads to a higher

level of social impact. Thus, we are able to test the assumption that open access monographs' availability beyond academic institutions leads to more usage by non-academics. However, other influences may affect usage. Differences in available infrastructure – the digital divide – is an example. Another possible factor is the dissemination platform: is it able to reach non-academics? Furthermore, aspects of the books such as language and subject play an important role. Any conclusion about the social impact of open access monographs based on usage data must account for these factors.

Besides usage data, other indicators of social and academic impact are also available. In the realm of journal articles, the number of citations is the most-used metric to assess academic impact. For monographs, citations are more problematic, which has been discussed in chapter 9. Investigating citation data for books is hampered by a lower availability of indexation services. Another challenging issue is the slower pace of citations, leading to a “citation window” of at least six to eight years. The third factor might be the difference in citation culture between scholarly disciplines. Lastly, in some fields of HSS, writing in English is not always the norm; this is problematic when citation indexes might be biased toward Anglo-Saxon regions (Nederhof, 2006). As is the case with usage data, any conclusion about the academic impact must take into account the special circumstances around open access monographs.

An indication of social impact might be found using altmetrics. Altmetrics share much characteristics with usage data. Instead of counting activities from infrastructure that is directly connected to documents, the usage of a broad range of social media and other online outlets is measured. As is the case with online book platforms, some outlets are more strongly directed towards academic users, while others are more open to everybody. For instance, online reference managers such as Mendeley or specialised websites such as ResearchGate are far more used by academics, while platforms such as Twitter or Facebook have a more diverse user base. On top of this, the different altmetrics outlets are also aligned differently to document types. Hammarfelt (2014) concludes that Mendeley is the best altmetrics outlet for humanities articles, while books are mostly mentioned on Twitter. In conclusion, Twitter is most likely to be used by the general public and mentions books most often. For that reason, the number of tweets is used as an indicator of social impact in chapter 9.

The next section discusses several examinations of the impact of open access on academics and non-academics.

10.5.2 Indications of impact

The question examined in chapter 7 is how to provide quantitative evidence of both academic and social impact of HSS research. The use of bibliometric data for monographs is problematic and the humanities and social sciences tend to place more emphasis on the societal impact of the results. Delivering evidence of impact depends for a large part on either self-reporting or in-depth discussion with stakeholders. Both methods are labour-intensive and susceptible to bias. Here, taking advantage of usage data might help to display another aspect: interaction with published results. Like altmetrics, the usage data is the direct result of online interaction, and the large number of data points enables the creation of sophisticated reports.

The usage data contains information about the organisation through which the reader accesses the web. By determining the type of organisation and the country of origin it is possible to assess the impact of the books, both in academia and beyond. The methods – tested on the OAPEN Library in 2011 – helps to uncover stakeholders, who may not always be known beforehand. Over 27% of the data is directly linked to academic users. In contrast, the usage linked directly to other “professional” users is less than 5%. The remaining 67% cannot be directly ascribed to the general public. The type of provider is a commercial Internet Service Provider (ISP), making it impossible to determine what organisation – if any – the reader is associated with.

In order to better categorize this large group of readers, I combined the available information about the country of origin with the state of intranet infrastructure. By using a fairly strict threshold, countries were grouped in those with a highly developed internet infrastructure, and those without. I assume that readers from a country with a highly developed internet infrastructure who download monographs out of a professional interest are more likely to use their organisation's internet infrastructure instead of an ISP. Thus, readers based in countries with a highly developed internet infrastructure that use an ISP to access the monographs, are more likely to be part of the general public. In this way, the large group of uncategorized users – 67% – can be classified. The smaller half of this group is still not categorizable, but the other half might be part of the general public in the wealthier countries of the world.

Apart from using provider types as proxies for users, the influence of scholarly discipline was analysed by looking at the differences in usage for humanities and social sciences books. Also, the differences in geographical impact of books in English versus books in Dutch are quite visible.

The question I examine is not whether more people are interacting with the monographs, the question is what kind of people are using the open access books. Usage should always be evaluated within the context of the platform. For instance, measuring usage of an academic library will not lead to finding many non-academic readers. The OAPEN Library is freely accessible and has taken several measures to make its content widely used, which might help to attract many different users.

When the results of chapter 4 are considered, we see that three-quarters of usage stems from direct access: incorporation into other systems than the OAPEN Library interface. The usage percentage from academic providers is less than 20%, while the usage through ISPs operating in countries with a highly developed internet infrastructure is 50%. Both the results of chapter 4 and of chapter 7 point to a relative low usage directly linked to academic institutions. The results of chapter 4 seem to suggest that other platforms than the OAPEN Library incorporate descriptions of the books. However, a large percentage of those platforms are not directly linked to an academic institution.

Thus, given these results it is feasible that the OAPEN Library's contents are available to readers beyond academic institutions in the "global north". The percentages directly linked to readers with a professional interest – those linked to government, non-profit or business organisations – are invariably low. And the largest single category consists of internet providers that have – at the very least – a possible link to readers that have downloaded the books for other than professional reasons. Returning to assessment within the context of a platform, its potential reach is wider than academic institutions alone. Consequently, the books available at an open access dissemination platform stand a good chance of reaching a wider audience. The percentage of monographs that are downloaded frequently and by other categories than academic institutions alone, are an indication of social impact.

Social impact is not restricted to the "global north"; does open access help to bridge the digital divide between those living in the richest countries and those in other parts of the world? Chapter 8 surveys whether open access enhances the use in the developing countries. In other words: does open access help to overcome the inequality in access to the internet, both in a technical sense and in lack of knowledge to optimally use the available resources? To test this, the usage data of the books in the experiment performed by Snijder (2010) were combined with geographical user data.

During the experiment – run in 2009 – several sets of monographs were made freely available. Another set of books was used as a control group.

The data was gathered from the Google Books platform; access to the books was strictly controlled for the experiment. This platform was ideally suited for this type of experiments: while all books on the Google Books platform were fully indexed by the Google search engine, it allowed publishers to decide what percentage of the book's contents were freely available. Thus, some books could be fully read online – 100% of the content available – and the control books showed no more than 10% of the text. To remove bias, the sets were carefully set up, based on subject; type of work; expected sales and publication date. The analysis in chapter 8 is based on 180 English language monographs.

Even when using a platform that is part of a globally used search engine, the digital divide between developed and developing countries is clearly visible: only 30% of the usage comes from developing countries. Further analysis of the differences between the usage of the open access books versus the closed access books revealed a more positive outcome. When reviewing the usage from developing countries and developed countries, the relative usage of open access monographs by developing countries was higher compared to the usage of the books that were not completely available. This is an indication of social impact: more usage of open access monographs by those in a disadvantaged position.

Before, I examined the possible influence of the collection's geographical focus and usage from the same region. While it might be a factor contributing to the lower usage from developing countries, the setup of the experimental and the control set of books helped to evenly spread the subjects. Additionally, the differences in internet infrastructure will have played a role in access and – in this case – the positive influence of open access is visible.

So far, I discussed usage as a means to measure academic and social impact of open access monographs. When the users are categorized by organisation type, academic users are the largest group. However, the combined download figures from academic organisations amount to roughly 20% of all downloads. In other words: it is possible to show the academic impact of open access monographs, but based on this data it is hard to conclude that open access enhances usage among academics. When I look at social impact, the results point toward increased usage by those who normally face additional challenges to access scholarly books: non-academics in the “global north” and those living and working in developing countries.

In order to answer the question whether open access has a positive influence in academia, I turned to another measurement: the number of citations. Many open access advocates have discussed the positive influence

on citations – seen by many as a major indicator of academic impact. Can a ‘citation advantage’ for open access monographs be found? This has been investigated numerous times for journal articles, but scarcely for books.

The research of chapter 9 dealt with these aspects in several ways. Instead of relying on a citation index, I used the Google Scholar platform. Secondly, to account for the “citation window”, the examined books were published at least five years before the date of obtaining the data. The differences in disciplines – and languages – have been dealt with in several ways. To maintain a balanced division of subjects and languages, I used the same sets of books as in Snijder (2010). Furthermore, I studied the influence of books in humanities versus other disciplines, plus additional testing on several groups of books on more specialised subjects.

To examine social impact, I used altmetrics. As mentioned before, some altmetrics sources are geared towards academic users, while others target a more diverse audience. Reference managers such as Mendeley are strongly related to academic use, while platforms like Facebook or Twitter are used by all types of internet users. In this research, I selected the altmetrics platform that performs best on monographs and is mostly connected to the general public: Twitter.

Given these preparations and choices, is it possible to establish whether open access has a positive effect on the number of citations? Also, does open access lead to more uptake by the general public? Looking at citations, the results are more or less in line with the literature on journal articles: a small but statistically significant positive effect of open access on the number of citations, even when the analysis takes into account the influence of language and subject. For tweets, the situation is slightly different. In the same way as citations, the average number of tweets about open access monographs is larger than the number of tweets about closed access books. However, the difference is not statistically significant. Lastly, little overlap exists between Twitter usage and citation behaviour. Thus, open access does not affect Twitter mentions in the same way as citations.

If citations are an indication of scholarly impact, the results point to positive influence of open access. However, the influence is not very large. A possible explanation might be found in the fact that a large proportion of researchers who are interested in the books in the data set, are in the position to access its contents anyway. This is supported by the outcomes of in chapter 8: in 2009, over 70% of usage of the English language books – which have a more global audience compared to the Dutch language books – was connected to the richest countries. Presumably, readers working in

those countries have a far better chance to view the books' contents, either through a library or by buying a copy.

Does this mean that the higher uptake of open access books – see Emery *et al.* (2017); Ferwerda *et al.* (2013); Snijder (2010) – can be mainly attributed to non-academics? Given the results of chapter 6, where roughly one-third of the usage has a larger chance to be associated with the general public, this might seem plausible. However, this conclusion is supported by indirect evidence. First of all, the data set of chapter 8 is based on 400 books. Whether this set is large enough to warrant such broad conclusions is questionable. Secondly, if I assume that the Twitter usage in chapter 9 indicates interest by the general public, the lack of statistically significant evidence is problematic. And lastly, I have discussed the differences in platforms. The users of the Google Book platform might differ significantly from the users of the OAPEN Library, and any conclusion spanning multiple platforms should be backed by solid evidence.

Direct evidence of the societal impact of open access monographs beyond the downloads of businesses, governmental organisations and non-profit organisation is not easy to obtain. Likewise, knowing what usage is related to the general public – which is by definition not affiliated to a specific type of organisation – is also problematic. Compared to journal articles, the available research data is still scarce. Therefore, more data is needed to provide more definitive answers, especially usage data and data about the collections of other open access book platforms. This will enable us to compare the effects of the identified factors on platforms with other collections and affordances: what are the effects on usage, citations or altmetrics? Hopefully, my research marks the start of more investigations.

10.6 Concluding remarks: factors affecting usage and the impact of open access

The introduction states that the level of open access monographs usage is primarily determined by book-related factors such as language and scholarly field or the configuration of dissemination platforms. The results show that these factors indeed affect the usage. Another factor is the level of trust in the content on offer. Contrary to expectations from several open access advocates, open licenses do not affect the level of usage. Furthermore, open access does not lead to more sales of monographs, yet it enhances usage in developing countries and the number of citations.

Most experiments in this dissertation have involved the collection of the OAPEN Library; a diverse set of books spanning multiple disciplines and languages. Therefore, it was relatively easy to measure how subject as proxy for scholarly field and language play an important role. For instance, while the topic of migration is not only discussed in academic circles but also in most newspapers, the audience for Sumerian spells¹ might be smaller. Usage is also connected to the geographical location of the readers: academic books discussing a certain part of the world tend to be read more by those who come from the same region. The usage of monographs written in other languages than English is also affected by geographic factors: books in German are more downloaded in German-speaking countries; the usage of Dutch language books is highest in The Netherlands.

The role of subject and language was to be expected. Furthermore, it is obvious that online dissemination is affected by the infrastructure that supports it. This has been clearly visible in the digital divide between rich and less well-off countries. Another aspect of online dissemination infrastructure is its interconnectivity: how well does one source integrate into another platform? The fact that the majority of the OAPEN Library downloads does not involve the front end can be seen as an illustration of the immersion into other systems.

Whether the technical abilities of dissemination platforms such as the OAPEN Library or the Directory of Open Access Books are used depends on a far less obvious factor: trust. Making a book available online does not automatically lead to optimal usage. Most people rely on filtering mechanisms to separate the wheat from the chaff. These mechanisms may include library catalogues, mentions on social media, specialised websites or blogs and many more possibilities. Additionally, the “filters” may rely on other sources: for instance, libraries might employ content aggregators.

In short, whether an open access monograph – or a platform that disseminates open access monographs – is accepted, depends on a conscious decision, not solely on an automated process. This is illustrated by the added usage from inclusion into the Directory of Open Access Books, but also by the inclusion of the contents of the OAPEN Library into other systems.

Ultimately, the decision to use an open access book platform is based on trust. Trust and the notion of quality are closely connected: when the books on offer are of sufficient quality, the prospective readers – or aggregators – will take action to obtain one or more books. As it is unlikely that each

1 Schramm, W. (2008). *Ein Compendium sumerisch-akkadischer Beschwörungen*. Universitätsverlag Göttingen. Retrieved from <http://www.oapen.org/record/610352>

book on the open access book platform will be vetted before downloading, the prospective readers – or the aggregators acting on their behalf – must assume that the offering is of sufficient quality. In other words, the readers must put their trust in the choices made by the platform.

Subject, language, infrastructure and trust are all influences that shape the usage of open access monographs. Other factors are not as important: licenses and the effects on sales. Licenses are seen as an important part of open access: the ability for readers to reuse the content has been described explicitly in the BOAI (Chan *et al.*, 2002). Given the emphasis on reuse, it was reasonable to expect more usage of monographs made available under a licence that actually permits it. However, whether an open access monograph licence only permits reading and downloading for personal use or enables content-sharing did not matter. Thus, the influence of licenses on usage is negligible.

The conclusion that open access does not affect the sales of monographs is not very surprising. I have been involved – directly and indirectly – in several experiments to measure the effect of open access on monograph sales (Collins & Milloy, 2016; Ferwerda *et al.*, 2013; Snijder, 2010; SNSF, 2015). In contrast to chapter 3, these experiments are based on a careful selection of monographs: an experimental set of titles that are published in open access, and a control group consisting of comparable books. None of these experiments resulted in a significant increase or decrease of the number of copies sold for the set of open access monographs.

Open access to monographs leads however to more usage in developing countries, a positive result. One of the goals of open access is enhancing the usage by those who would otherwise not be able to read scholarly output. Here, this goal has been achieved, albeit on a small scale. Another often-used benchmark in the realm of journal articles is the “citation advantage” of open access publications. For monographs, I was able to demonstrate a slight citation advantage.

To recapitulate, while open access monographs dissemination is only possible by removing paywalls, the level of usage is primarily determined by language, subject, infrastructure and trust. Given these influences, open access enhances usage in developing countries and the number of citations.

10.7 Practical implications and further research

What are the practical implications of these results? In my opinion, an open access monographs platform should focus on trust. After all, trust is

the most important aspect: any platform can only be successful if people want to use it. Transparency about the selection criteria of the collection helps prospective readers and aggregators to determine whether the books on offer are of interest to them. In the case of OAPEN, the criteria (quality controlled monographs) are listed on the homepage, and the quality control process of the publishers is described.² The Directory of Open Access Books has adopted a similar policy.³

When a reader or an aggregator wants to use the content, the platform should make it easy to connect. We have seen before that the OAPEN Library is used via several channels: not just as an online public access catalogue, but also as a web based database that can be integrated into a larger collection. To ensure technical integration, the platform should offer its metadata based on standards that are used by the reader or aggregator. For instance, OAPEN supports aggregators with metadata feeds based on ONIX – a standard used in the publishing industry – and MARC21 – a library standard. Readers who are interested in a single title can download metadata in RIS format – to be used in citation managers – or use a widget to share the description via social media and mail. Connecting with readers or aggregators ought to go beyond technical measures. In the case of the OAPEN Library and DOAB, this is translated into agreements with commercial and non-commercial aggregators and by using social media to connect to individuals.

Language has proven to be an important influence on usage. Furthermore, the bulk of the usage of the OAPEN Library so far stems from the “global north”. To extend the usage to the rest of the world, it might be useful to add monographs in Spanish and Portuguese to the collection – languages that are spoken in Latin-America. Also, a larger collection in French might be more attractive to the French-speaking countries in Africa.

The results so far are a good start towards understanding the effects of open access on monographs and the factors affecting usage of open access monographs. However, further research could help to deepen our understanding. The first research question would be the identification of usage by the general public. In my research, recognizing members of the public was based on eliminating possible organisational ties. Research on this topic should take into account privacy considerations; this has also been discussed in chapter 6.

2 <http://oapen.org/content/peer-review-process-introduction>

3 <https://doabooks.org/doab?func=about&uiLanguage=en#purpose>

In this dissertation, most of the research has been carried out on the OAPEN Library platform. As more open access book platforms are emerging, it will be interesting to repeat some of the experiments on those platforms. What are the effects of differences in technical abilities and book collections? Comparing usage results of multiple platforms has its own challenges; the COUNTER Code of Practice (COUNTER Online Metrics, 2014) might be useful in this case.

The effects of open access on the usage originating from developing countries has been discussed in detail. However, the collection of titles in this investigation have been provided by publishers from the “global north”. If the collection of titles is enhanced with a sizable portion of titles from “global south” publishers, how would that affect the usage data? Does this lead to a higher percentage of usage from developing countries? Will the enhanced exposure be beneficial for authors?

I have deployed a clustering algorithm to find related books, based on usage by readers. The next phase would be to test several algorithms, in order to see if other procedures lead to comparable results. This will strengthen the claims of chapter 6. A related question is whether new algorithms lead to more fine-grained clusters.

Related to clustering algorithms, using text mining techniques to extract subjects from books might lead to new possibilities, for instance automatically clustering books based on distinctive words or word sequences and comparing these ‘subject clusters’ with the clusters of providers that were created for chapter 6.

Another possibility, based on the contents of the books, is to automatically define distinctive text segments, and searching whether they are used in newspapers, reports and other non-academic documents. This might help to determine the social impact of the monographs. The same technique could also be used as a service to readers, by searching for related academic open access documents in large databases such as BASE - Bielefeld Academic Search Engine.⁴

When the focus is widened beyond questions of usage, we might look at the role of paper books. Open access is inherently digital – based on online dissemination. Still, the role of paper books is not obsolete: the lack of influence of open access on sales of ‘traditional’ monographs points in that direction. Each publication form has its own merits, but it would be interesting to investigate whether the ideal of world wide free dissemination of knowledge can be combined with the affordances of paper publications.

4 <https://www.base-search.net/about/en/>

The success of this approach will depend on the stakeholders in scholarly publication.

Changes in online dissemination and the variations in stakeholder roles were already briefly discussed in section 10.4. The effects of this transition merit further research: if publishers continue to build online libraries, and academic libraries keep enlarging their publishing role, how will this affect scholarly communication?

In conclusion, the research on the dissemination of knowledge through open access monographs is far from finished. We have barely started.