

Books in a bubble: assessing the OAPEN Library collection

Ronald Snijder^(a)

a) OAPEN Foundation, Netherlands, <https://orcid.org/0000-0001-9260-4941>

Contact: Ronald Snijder, R.snijder@oapen.org

Received: 12 August 2022; **Accepted:** 21 December 2022; **First Published:** 15 May 2023

ABSTRACT

Open access infrastructure for books is becoming more mature, and it is being used by an increasing number of people. The growing importance of open access infrastructure leads to more interest in sustainability, governance and impact assessment. The assessment of the OAPEN Library – containing over 20,000 titles – fits within this trend. How well does the collections meets the needs of its stakeholders: readers, libraries, funders and publishers?

The composition of the collection is measured using subject and language. Both dissemination and the content-related aspects are paired to the number of publications. The average number of downloads per title is relatively similar for all subjects. However, the mean downloads of titles in English is roughly twice as much compared to German and the other languages. Combining subjects and languages shows that the dissemination of books in languages other than English is less predictable. This assessment has illustrated the composition of the collection and how its readers make use of it. The visualisation helped to tell a complicated story in a simple way; a powerful instrument to guide the further development of this open access infrastructure.

KEYWORDS

Open access; OAPEN Library; Library assessment; Books; Book impact.

1. Introduction

Open access infrastructure for books is becoming more mature, and it is being used by an increasing number of people. The growing importance of open access infrastructure leads to more interest in sustainability, governance and impact assessment. For this paper, we will assess the OAPEN Library. In the spring of 2022, it passed the milestone of 20,000 titles. This was a good moment to evaluate the core asset of the OAPEN Library: its collection.

The OAPEN Library contains freely accessible books and chapters, all of which have undergone external peer review. In other words, it functions as an academic library and in our assessment we should treat it as such. However, it is an online library and the limitations of physical books do not play a part. Shelf space is not to be considered.

More important is the question of how well the collection meets the needs of its users. The OAPEN Library sees global usage; the collection reflects this by offering titles in over 50 languages. The collection is not focused on a specific subject area, but the choice of medium – books and chapters, not journals and articles – is more strongly associated with the humanities and social sciences. It does not track its users, but the supporters of the OAPEN Libraries are globally distributed academic institutions, scientific and scholarly funders and publishers. An assessment of the OAPEN Library should therefore take into account the diversity of languages, subjects and stakeholders.

The following section describes the history of the OAPEN Library and its closely related platform: the Directory of Open Access Books.

2. The OAPEN Library and DOAB

The OAPEN Library was officially launched in 2010 and has been set up to host and disseminate open access books and chapters (Ogg 2010). Together with publishers OAPEN is building a quality-controlled collection; OAPEN also provides services for publishers, libraries, and research funders in the areas of dissemination, quality assurance, reporting and digital preservation. Furthermore, OAPEN collaborates with research funders, allowing the deposition of books that have received financial support from several funders. These deposited titles also must have been subject to an external quality control. In 2013 OAPEN improved its reporting by providing COUNTER-conformant usage statistics in collaboration with IRUS-UK (IRUS-UK 2018).

The OAPEN Library is built using open-source software and all its metadata has been made available under a CC0 licence. It is not optimized to keep visitors as long as possible on the site to show them advertisements or sell them as many things as possible. Instead, it maximises the dissemination of books and chapters using as many venues as possible. Books might be discovered on the OAPEN Library, or through the Directory of Open Access Books, library catalogues, Google Scholar, OpenAIRE, Twitter, a blog or by any other means.

When OAPEN was in its project phase, discussions on setting up an additional services were taking place. Based on the Directory of Open Access Journals (DOAJ) – which was launched in 2003 – the idea of an directory listing open access books was developed. This led to the development of the Directory of Open Access Books (DOAB). It would differ from the OAPEN Library in several aspects. Firstly, it would not host the full text of books but provide a searchable index

with links to the publications at the publisher's website or repository. Secondly, while OAPEN allows books that may be downloaded for private use only, all books listed IN DOAB should be published under an open licence, allowing re-use of the publications. And lastly, the service would be free for publishers that were approved to submit their publications.

DOAB was officially launched in 2013, as a joint service of OAPEN, OpenEdition, CNRS and Aix-Marseille Université, provided by the DOAB Foundation (Whitford 2014). Since then it has become a global focal point for open access books. The growth of the collection – in December 2022 it contained over 63,000 title descriptions – was also aided by other organisations that have become part of the “DOAB Trusted Platform Network”: OpenEdition; Project MUSE; SciELO Books (DOAB 2021). The fourth member of the DOAB Trusted Platform Network is the OAPEN Library. In 2020 and 2021, both the OAPEN Library and the Directory of Open Access Books have been migrated to the DSpace 6 platform.

In the next section, we will discuss the literature on library and collection assessments.

3. Literature review

Assessments of libraries and their collections are taking place within a certain context. For instance, much of the research literature discusses the distinction between print holdings and electronic holdings. In libraries containing curated print collections and a more general digital collection it is important to measure the usage of both collections, which is called “book impact”. In several cases, print holdings are more strongly correlated with book impact, while this correlation is less clear for electronic holdings. (Borin and Yi 2008) describe the changing environment in which libraries find themselves, where electronic titles play a more important role, leading to different demands on how the library must be operated. An assessment model must be flexible enough to function. In a subsequent article, (Borin and Yi 2011) describe the deployment of their model. Measuring the impact of a digital collection is challenging. (Perrin et al. 2017) describe the complexities of understanding the data generated by Google Analytics of a DSpace environment.

Other authors focus more on the stronger and weaker points of the library collection; an example is the work by (Hyödynmaa, Ahlholm-Kannisto, and Nurminen 2010) on collection mapping in Tampere University Library. At the George Mason University, (Kelly 2014) combines a multi-tool analysis with a systematic, subject-by-subject approach to the collection. And finally, (Lantzy, Matlin, and Opdahl 2020) describe a continuous library assessment with recurring cycles.

What is the result of these library assessments? According to (Folb, Wessel, and Czechowski 2011), the patrons of the University of Pittsburgh are accepting e-books as an alternative to print books. (Lamothe 2010) correlates usage of electronic books to the amount of curation applied to the collection and the collection size. The next step is to make the research results easy to understand and communicate. Visualisation can be a useful tool to convey complex ideas. (Chen 2017) describes several visualisation techniques, and (Finch and Flenner 2016) introduce the bubble charts that will be used in this paper.

We could argue that libraries are much more than their collections, and that the level of openness is an important precursor. If the library has a function beyond its parent organisation, it might

serve an even larger community. (Brookes-Kenworthy et al. 2019) looked into the relationship between academic library access policies and institutional positions on open access or open science publishing. Another aspect is policies towards unaffiliated user access, the so-called “walk-in users”, as observed by (Stregger and McLean 2021). In a similar vein, infrastructures like the OAPEN Library could also be assessed based on guidelines such as the Principles of Open Scholarly Infrastructure (Bilder, Lin, and Neylon 2020).

4. Data

In the previous section we have seen that library and collection assessments are complex and have to take into account the context of that specific library. Thus, what are the relevant factors for the OAPEN Library? Biodiversity¹ might be the most prominent aspect: the collection consists of publications from many academic publishers in multiple languages and is used by a global audience. Furthermore, the collection does not have a specific subject-oriented focus but the publication forms – books and chapters – are more used in the humanities and social sciences and less in the disciplines of science, technology and medicine.

Three types of stakeholders support the OAPEN Library: publishers, funders and libraries. Both publishers and funders contribute to the collection by making publications available. They will be interested in the dissemination of the books and chapters. For libraries, the composition of the collection will be paramount. How do the titles on offer fit within the information needs of their patrons?

The assessment of the OAPEN collection should consider these two aspects. The dissemination of books and chapters is measured through the number of downloads – based on COUNTER R5 conformant data. The composition of the collection is measured among two axes: subject and language. Both dissemination and the content-related aspects will be paired to the number of publications. The number of downloads are measured during a period of 12 months, ending March 2022.

The data set consists of 20,517 titles. For this assessment, the subject description and language of each title has been reduced to one major term. OAPEN uses the BIC Standard Subject Categories (Book Industry Communication 2010). The following subject listings have been used: “The arts”; “Economics, finance, business & management”; “Humanities”; “Language”; “Literature & literary studies”; “Society & social sciences” and all other topics are grouped under “Other subjects”. Languages have been grouped into: English; German; Other. During the period April 2021 until March 2022, the total usage of the collection consisted of 11,876,706 downloads.

The data is freely accessible at <https://doi.org/10.5281/zenodo.6985305>.

The next section discusses the assessment of the OAPEN collection in more detail.

¹ “[T]he diversity of academic content, both at the national and international level is essential for preserving research in a wide range of global and local topics, studied from different epistemic and methodological approaches, inspired by various schools of thought and expressed in a variety of languages.”(Giménez Toledo et al. 2019)

5. Assessment of the OAPEN collection

In this paper the assessment of the OAPEN collection is based on several charts. All charts list three dimensions: the number of titles on the X-axis; the number of downloads on the Y-axis; and the size of the bubble's circumference displays the average number of downloads.

It is clear that the number of titles directly affects the number of downloads; larger groups of books and chapters will also accumulate more usage. Comparing the average number of downloads for the separate groups gives an indication of dissemination patterns: are certain sets of publications more 'popular' than others?

All charts are based on the distribution of titles within the complete collection. When we list the subjects and the languages in one table, several things are immediately visible. First of all, the dominance of English: 80% of the collection is written in this language. Secondly, the six subjects 'encompass' 76% of the collection, which means there is a considerable 'long tail' of titles discussing other topics, among them are science, technology and medicine.

Table 1 Percentages of subjects and languages

	<i>English</i>	<i>German</i>	<i>Other</i>	<i>Total</i>
Society & social sciences	23%	4%	2%	29%
Humanities	13%	2%	1%	16%
Economics, finance, business & management	7%	1%	0%	8%
Literature & literary studies	4%	1%	1%	7%
Language	3%	1%	1%	6%
The arts	5%	1%	0%	6%
Other subjects	24%	3%	1%	28%
Total	80%	12%	8%	100%

In the next subsections, the relations between these aspects and the dissemination are described in more detail. Dissemination is measured as the average number of downloads for a group of titles. Just as it is a simplification to reduce a book's subject to one main category, calculating the average number of downloads might hide large differences.

5.1 Collection by subject

Figure 1 displays the complete collection, sorted by subject. The large difference between “Social sciences” and “Humanities” compared to the other subjects is clearly visible, as is the large number of titles that are categorized under “Other subjects”. What is interesting to note is that the average number of downloads per title is relatively similar for all subjects: between 496 and 638. From a dissemination perspective we can assume that subject is not a large factor. To put it differently, the users of the OAPEN Library seem to be equally interested in all the subjects the collection has to offer.

However, language is another important factor affecting usage, which is depicted in the next subsection.

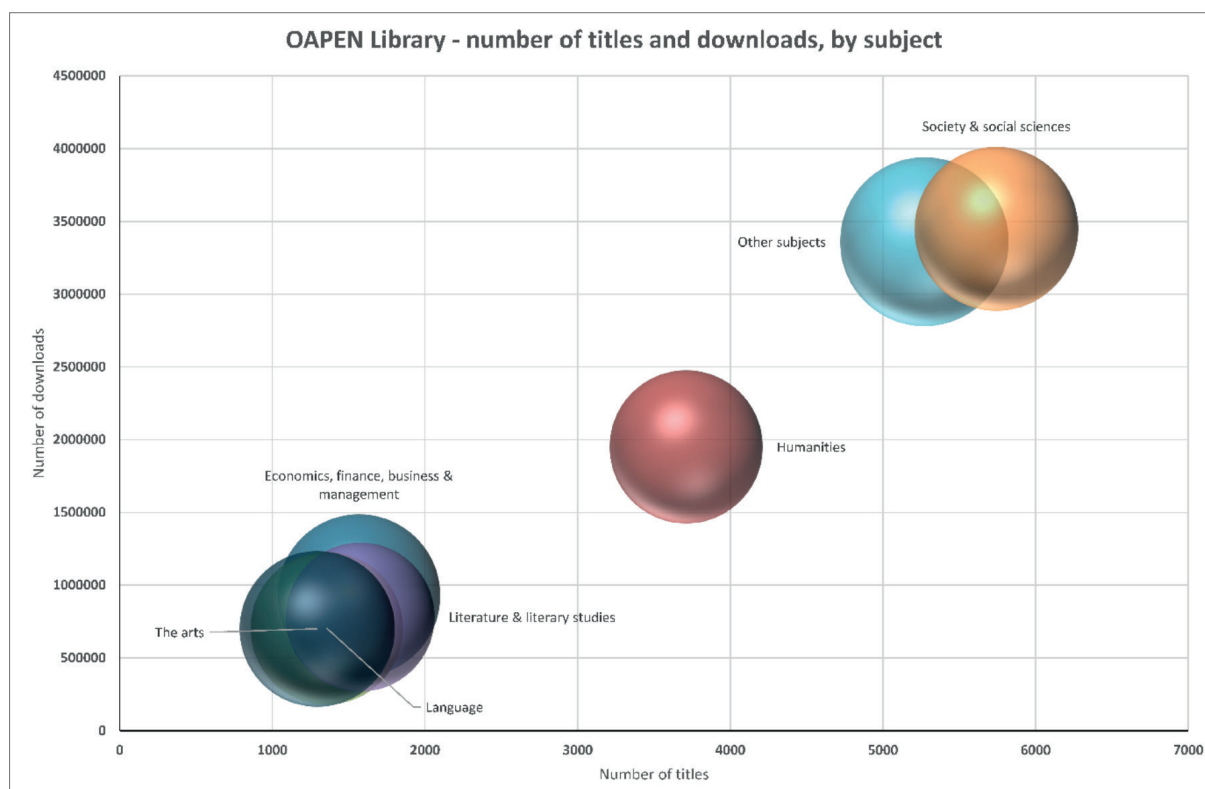


Figure 1. Number of titles and downloads, by subject.

5.2 Collection by language

The averages in Figure 2 differ quite dramatically. The mean downloads of titles in English is roughly twice as much compared to German and the other languages. And it is obvious when we look at the number of publications that English is the dominant language in the collection.

Here we also see that a larger number of titles does not always lead to higher average use. While the number of German-language books and chapters is twice as high as the number of titles in other languages, the average number of downloads of German titles is only 75% of the average for other languages.

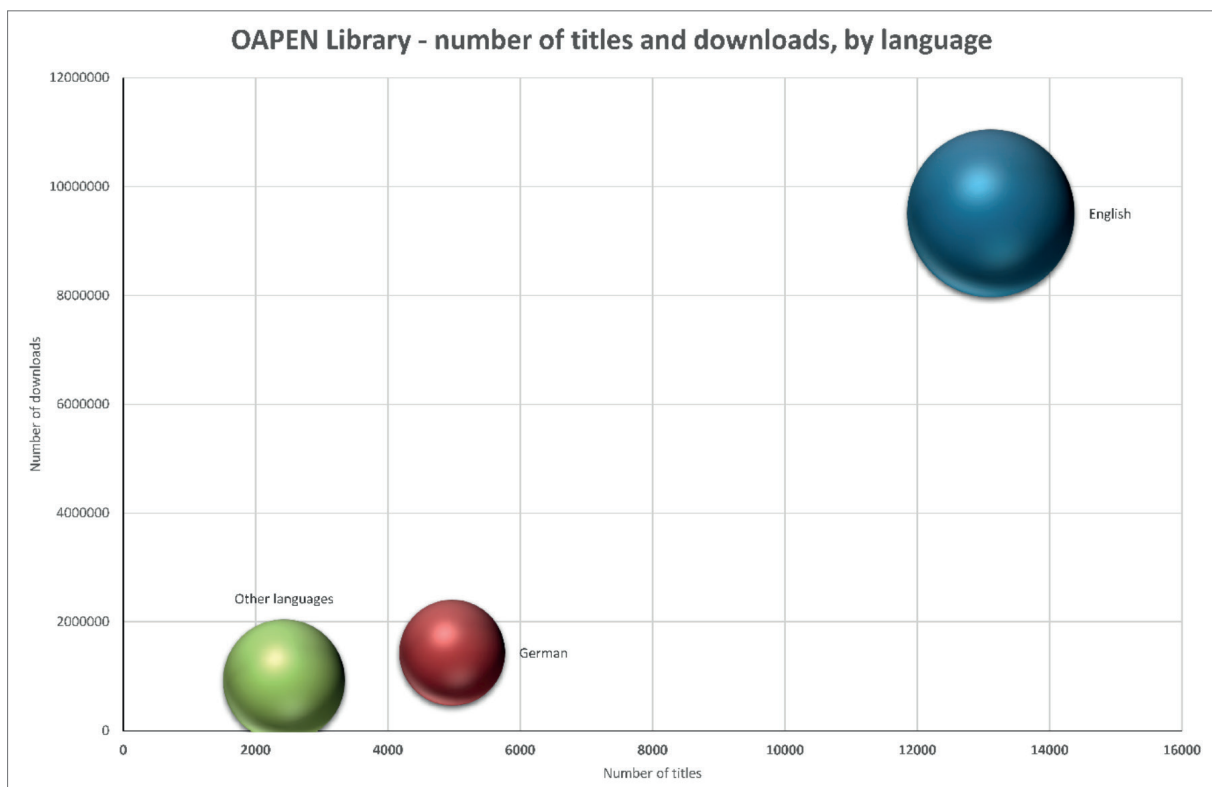


Figure 2. Number of titles and downloads, by language.

5.3 Subject and language combined

In this subsection, all combinations of subject and languages are illustrated. Within each of these sets, the number of books in English is always higher than the number of German-language titles or titles in other languages. Nevertheless, the dissemination of books in languages other than English tends to be less predictable. The average number of downloaded publications in other languages is much higher for “Humanities”, “Literature” and “the Arts”. In the other subjects, there is no such large difference.

5.3.1 Social sciences

The collection contains 5,743 publications on social sciences.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	3,500	2,763,290	790
German	1,448	468,247	323
Other	795	218,161	274

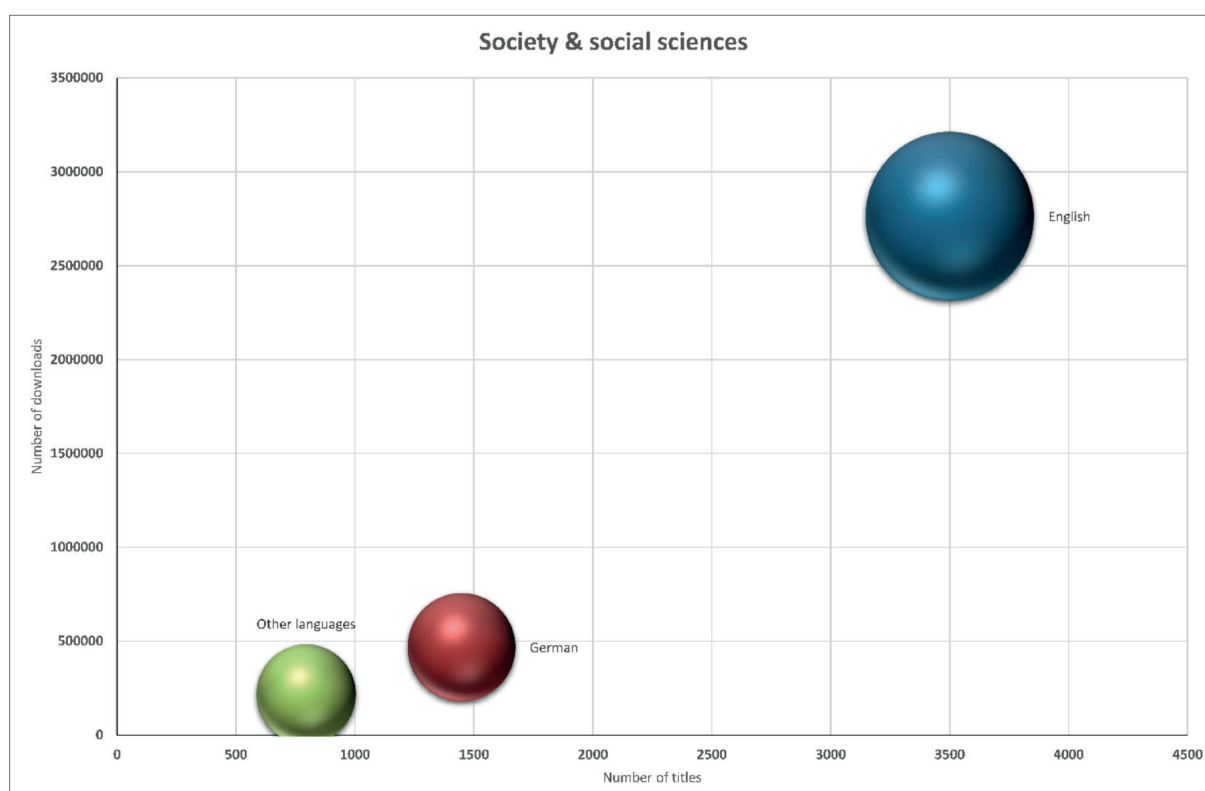


Figure 3. Social sciences.

5.3.2 Humanities

The collection contains 3,711 publications on humanities.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	2,443	1,583,925	648
German	787	199,119	253
Other	481	167,592	348

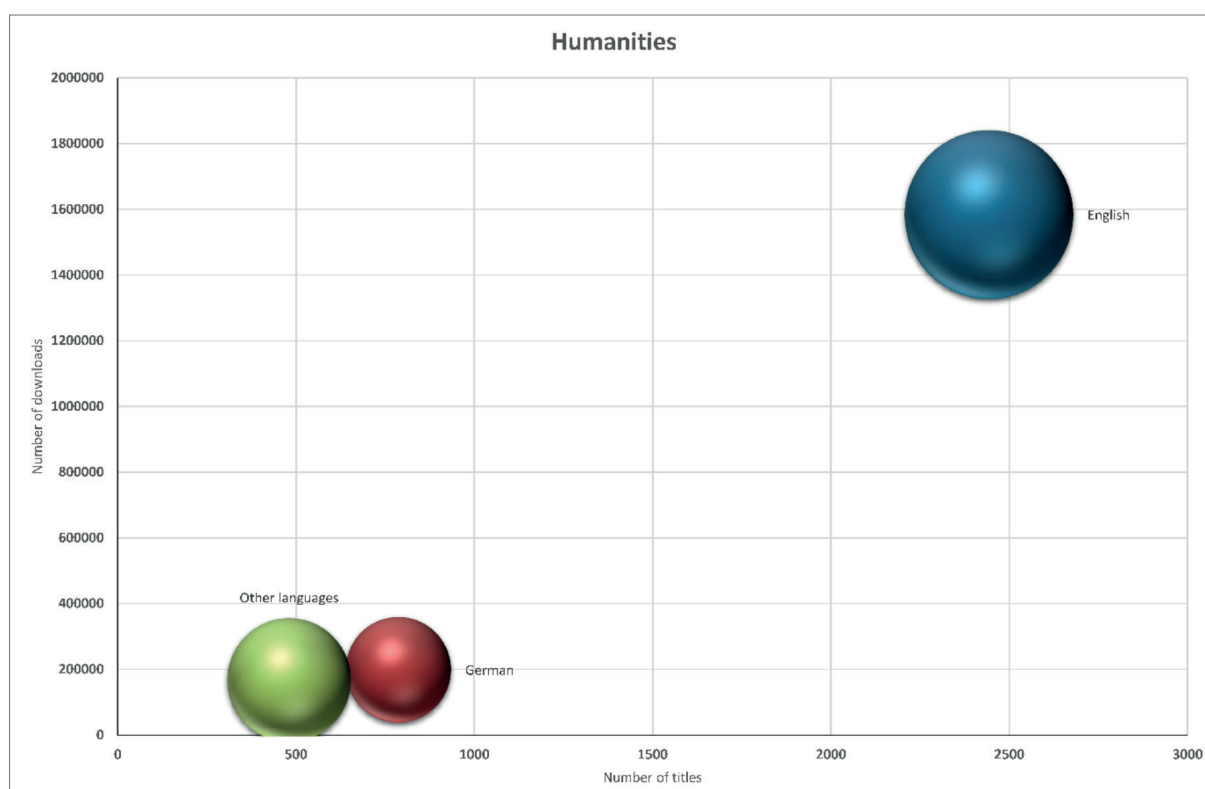


Figure 4. Humanities.

5.3.3 Economics

The collection contains 1,567 titles on economics.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	969	787,851	813
German	476	119,411	251
Other	122	21,437	176

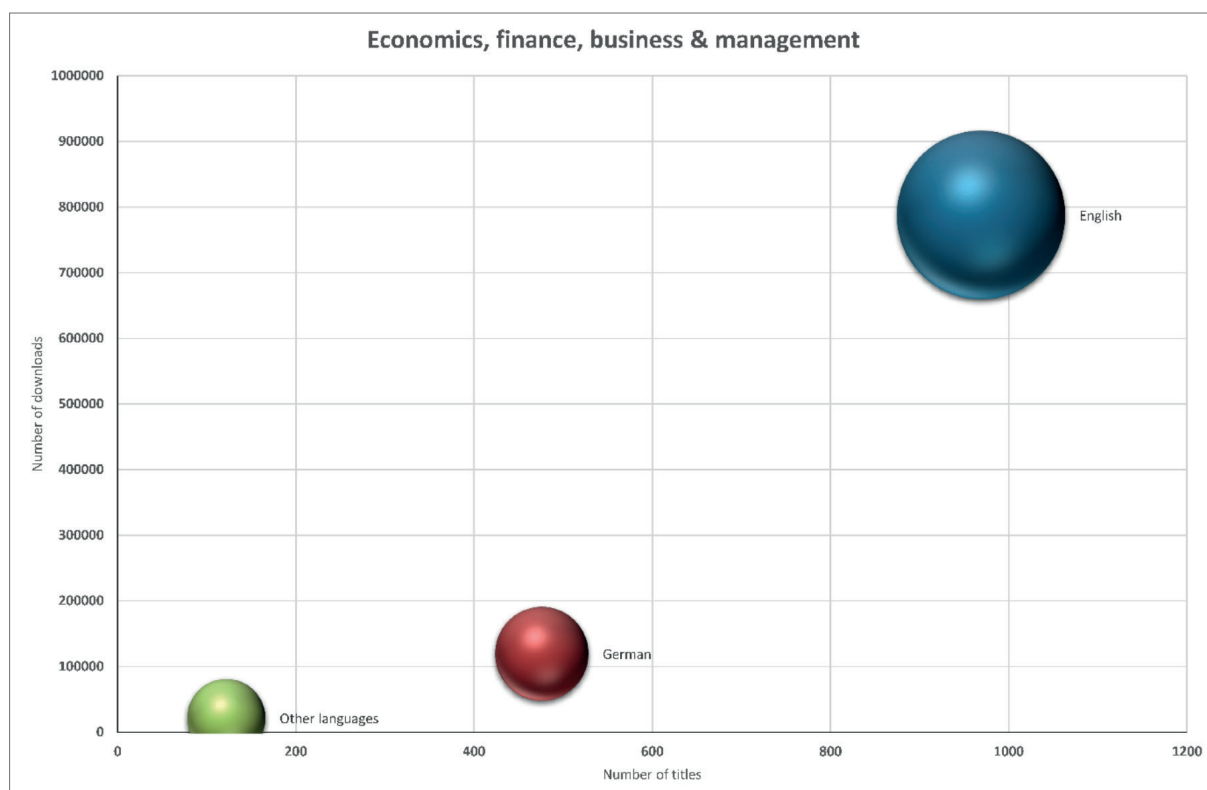


Figure 5. Economics, finance, business and management.

5.3.4 Literature

The collection contains 1,575 titles on literature.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	823	520,149	632
German	492	113,699	231
Other	260	147,371	567

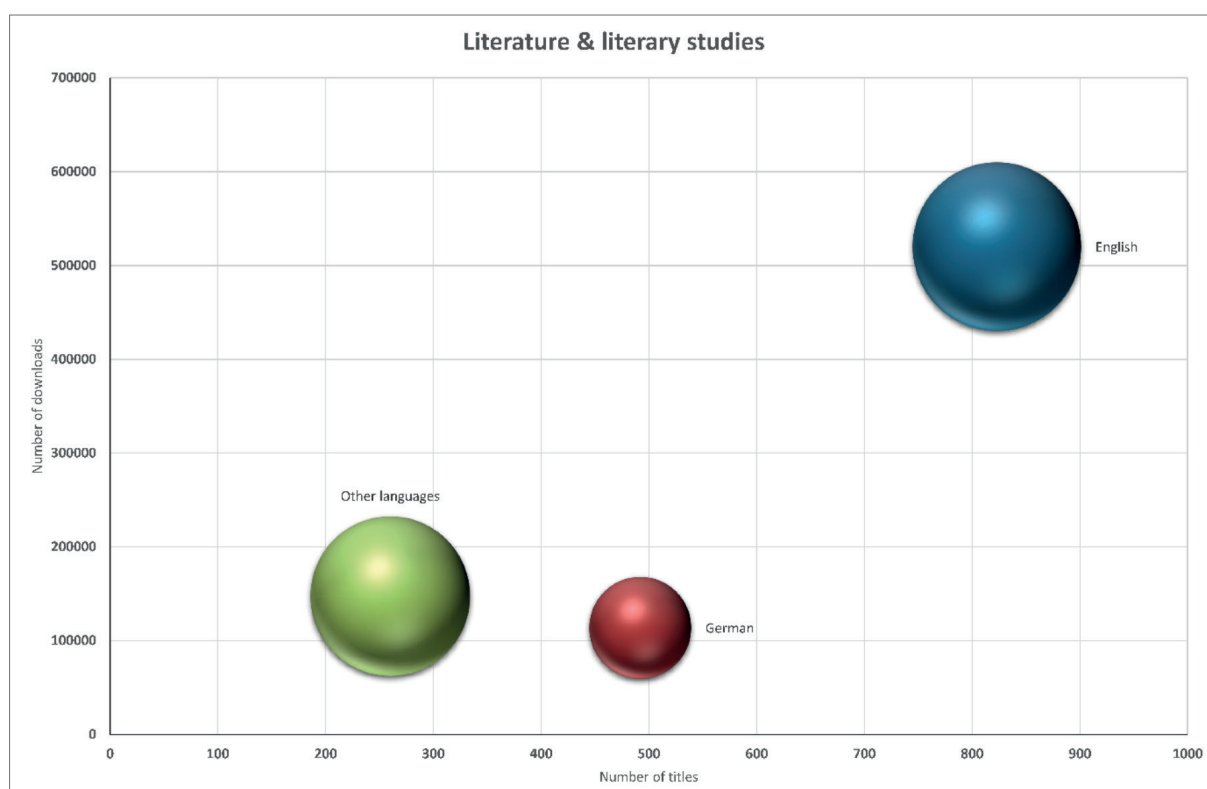


Figure 6. Literature and literary studies.

5.3.5 Language

The collection contains 1,356 titles on language.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	656	392,272	598
German	488	144,530	296
Other	212	167,841	792

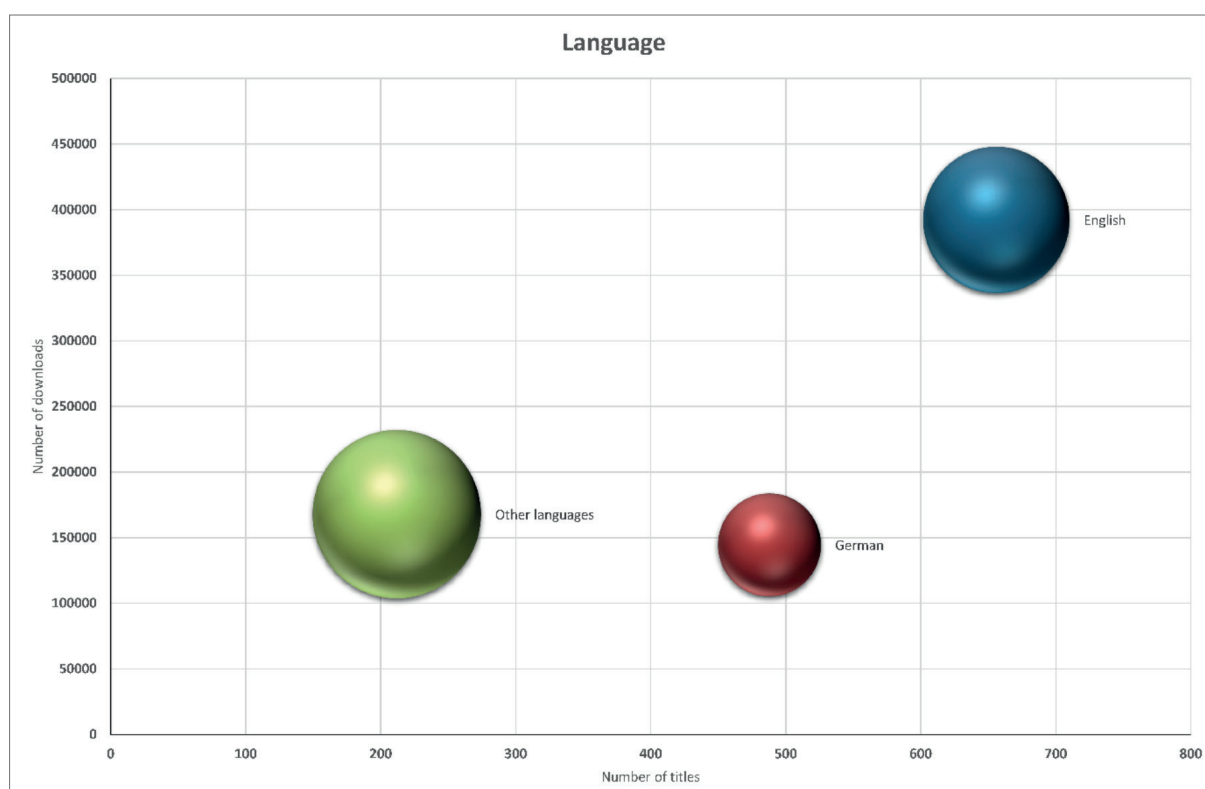


Figure 7. Language.

5.3.6 The Arts

The collection contains 1,294 titles on the arts.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	855	576,126	674
German	301	72,615	241
Other	138	51,814	375

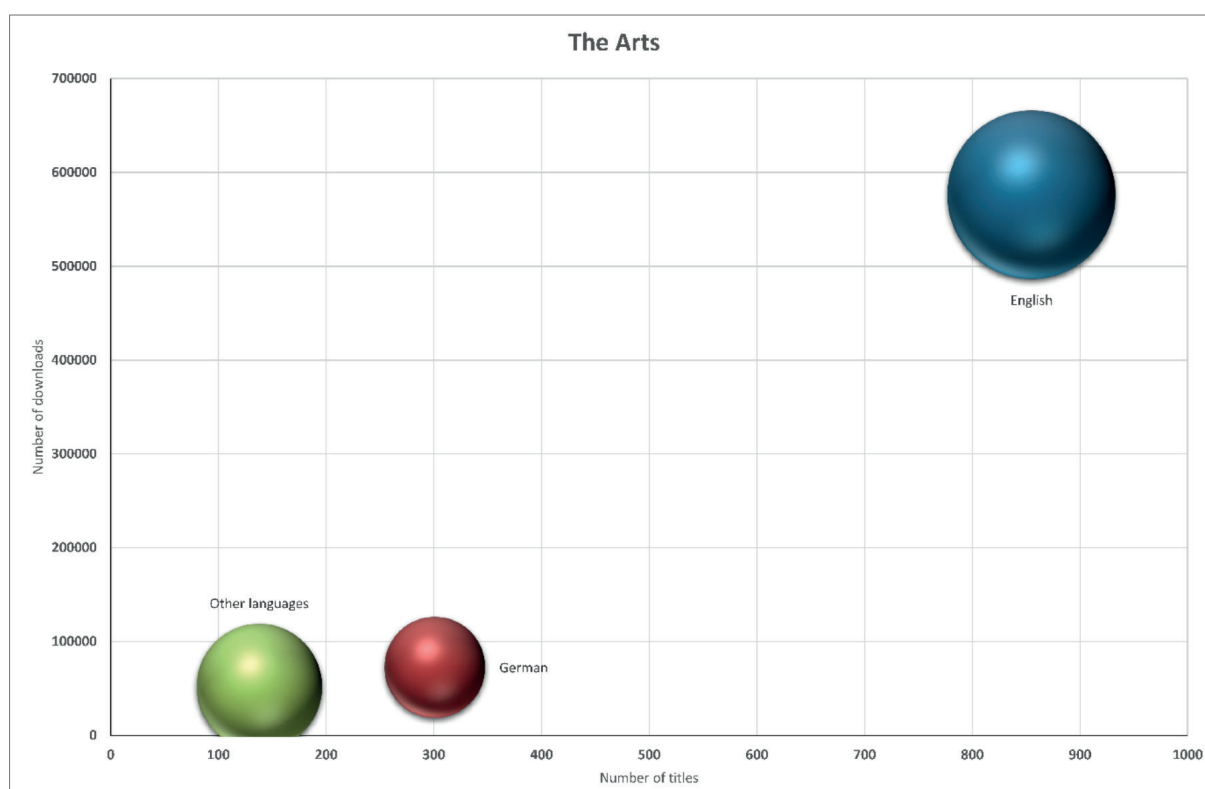


Figure 8. The arts.

5.3.7 Other subjects

The collection contains 5,271 titles on other subjects.

<i>Language</i>	<i>Number of titles</i>	<i>Total downloads</i>	<i>Average downloads</i>
English	3,868	2,887,631	747
German	979	317,828	325
Other	424	155,797	367

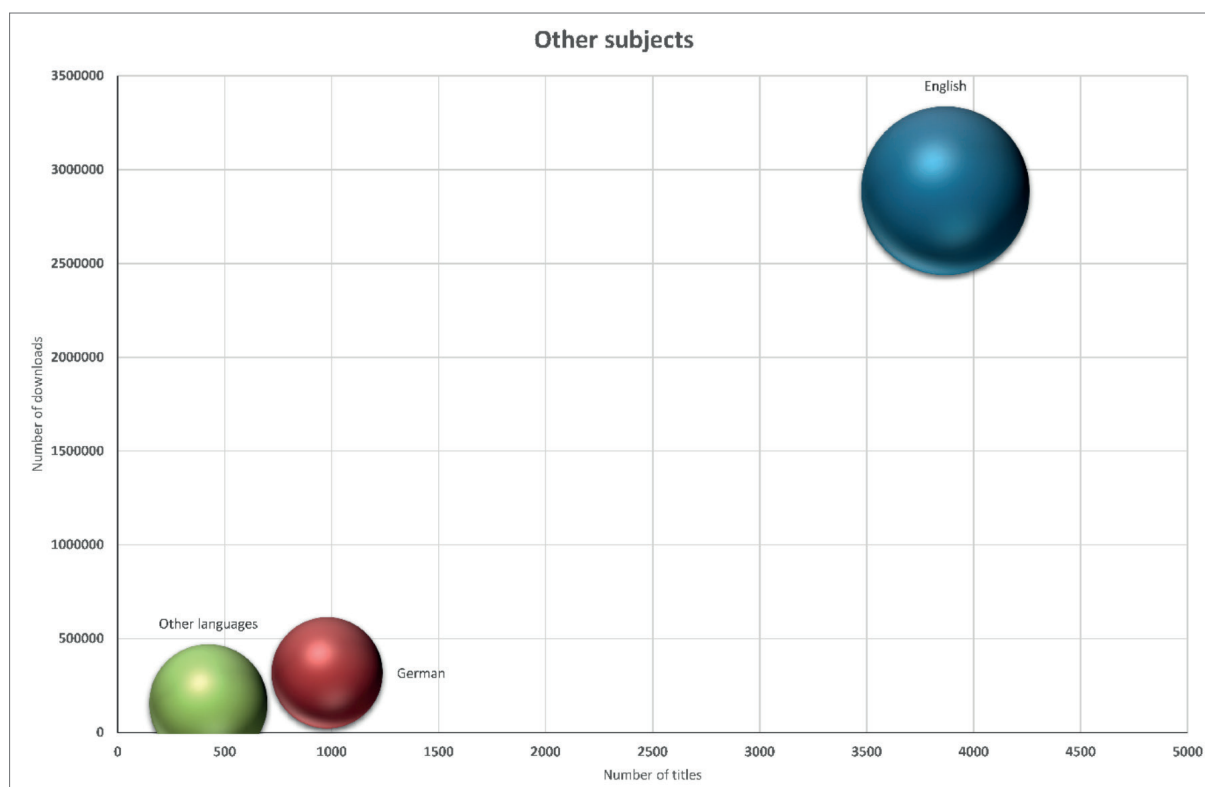


Figure 9. Other subjects.

6. Discussion

As we have seen in the literature review section, the assessment performed by a library should fit within its environment. Different circumstances lead to different questions. The OAPEN library operates in a different way from a more traditional library; it makes no distinction between members and visitors, as OAPEN Library content is freely available to everyone, but also functions as a supplier of an online collection for other libraries. This collection can be added completely or partially, leading to the question of the composition of the set of titles on offer. The assessment reflects this by focusing on subjects. However, within the humanities and social sciences, publication languages other than English are of more importance compared to the sciences. While most of the books and chapters are written in English, over 7,400 titles have been published in German or other languages.

Within the context of OAPEN, both publishers and funders are also important stakeholders. Publishers provide OAPEN with peer reviewed content. The collection is not subject based; any books and chapters can be added, provided that they have undergone an external quality assessment and are freely available to read and download. Both the publishers and the funders have an interest in how well the books and chapters are disseminated, and this is measured using COUNTER conformant download figures. Those numbers are affected by the language and the subject of the publications. That is why the combination of subject and language provides a better indication than the averages per subject alone.

The previous section displayed several patterns. Firstly, the dominance of English is visible in the number of titles available in the collection, and in the usage data. A larger number of titles will almost certainly lead to more downloads in total, but when we look at the average downloads per title, it is clear that English is more read by the users of the OAPEN Library. However, the impact of other languages varies strongly per subject. German is quite clearly the second largest language when we look at the number of publications. The total number of downloads for books and chapters in this language is consistently larger than the number of downloads for titles in other languages. Nevertheless, the average usage per title fluctuates. In the case of “Humanities”, “the Arts” and especially “Language” and “Literature”, the average number of downloads is higher for books in other languages. For these subjects, the collection of publications in other languages than English or German is smaller, but its book impact is larger.

This analysis did not take into account another important aspect of the usage: the regional diversity of the readers. We have seen that language has a large influence on the average number of downloads, especially when we compare English to other languages. Previous research on the international usage of the OAPEN Library has shown that readers from many countries have a strong preference for books written in their native or widely used languages (Snijder 2022). For instance, the ten most downloaded books in Chile were written in Spanish, while ten French language titles are most widely read in Cote d’Ivoire.

However, this also adds another layer of complexity to the assessment. Currently, the titles have been simplified by assigning them just one language and one subject. If we are to condense the geographical aspects of the readers, what choices need to be made? A possible option is to divide by the largest use per country. However, this would put much emphasis on large countries with a well-established internet infrastructure. Another possibility could be to create groups based on

continents. Of course, this will also hide away much of the diversity. Further discussion would be required to determine whether regional diversity of the readers should be added to these types of assessments and if so, how this should be implemented.

7. Conclusion

As can be seen from the results, even when the collection has been simplified to titles with one subject and one language, the usage patterns are not altogether straightforward. Comparing absolute download numbers are useless when the context – the subject and language – are not taken into account. Creating benchmarks to compare the book impact of a specific title is certainly a possibility, but only when the context has been taken into account.

Furthermore – and this is an important reason behind the continuous evaluation described by (Lantzy, Matlin, and Opdahl 2020) – this assessment is a “snapshot” depicting the situation at a certain time. Since the spring of 2022, the collection of the OAPEN Library has grown, not in the least with the addition of over 1,400 Italian-language titles. Obviously, a larger and more diverse collection will lead to different assessment results.

In conclusion, this assessment has vividly illustrated the composition of the collection and how its readers make use of it. The visualisation helped to tell a complicated story in a simple way; a powerful instrument to guide the further development of this open access infrastructure.

8. Acknowledgements

The author would like to thank the OAPEN colleagues for commenting on a previous version of this paper.

References

- Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. 2020. 'The Principles of Open Scholarly Infrastructure'. <https://doi.org/10.24343/C34W2H>.
- Book Industry Communication. 2010. 'BIC Standard Subject Categories – an Overview'. 2010. <http://www.bic.org.uk/7/BIC-Standard-Subject-Categories/>.
- Borin, Jacqueline, and Hua Yi. 2008. 'Indicators for Collection Evaluation: A New Dimensional Framework'. *Collection Building* 27 (4): 136–43. <https://doi.org/10.1108/01604950810913698>.
- . 2011. 'Assessing an Academic Library Collection through Capacity and Usage Indicators: Testing a Multi-dimensional Model'. *Collection Building* 30 (3): 120–25. <https://doi.org/10.1108/01604951111146956>.
- Brookes-Kenworthy, Chloe, Richard Hosking, Chun-Kai (Karl) Huang, Lucy Montgomery, Cameron Neylon, Alkim Ozaygen, and Katie Wilson. 2019. "Is the Library Open?": Correlating Unaffiliated Access to Academic Libraries with Open Access Support', September. <https://hcommons.org/deposits/item/hc:26879/>.
- Chen, Hsuanwei Michelle. 2017. 'Chapter 4. Real-World Uses for Information Visualization in Libraries'. *Library Technology Reports* 53 (3): 21–25. <https://www.journals.ala.org/index.php/ltr/article/view/6291>.
- DOAB. 2021. 'DOAB Trusted Platform Network'. 2021. <https://www.doabooks.org/en/publishers/doab-trusted-platform-network>.
- Finch, Jannette L., and Angela R. Flenner. 2016. 'Using Data Visualization to Examine an Academic Library Collection'. *College & Research Libraries* 77 (6): 765–78. <https://doi.org/10.5860/crl.77.6.765>.
- Folb, Barbara L., Charles B. Wessel, and Leslie J. Czechowski. 2011. 'Clinical and Academic Use of Electronic and Print Books: The Health Sciences Library System e-Book Study at the University of Pittsburgh'. *Journal of the Medical Library Association : JMLA* 99 (3): 218–28. <https://doi.org/10.3163/1536-5050.99.3.009>.
- Giménez Toledo, Elea, Emanuel Kulczycki, Janne Pölönen, and Gunnar Sivertsen. 2019. 'Bibliodiversity-What It Is and Why It Is Essential to Creating Situated Knowledge'. *Impact of Social Sciences Blog*. http://eprints.lse.ac.uk/103440/1/impactofsocialsciences_2019_12_05_bibliodiversity_what_it_is_and_why.pdf.
- Hyödynmaa, Merja, Anniita Ahlholm-Kannisto, and Hannele Nurminen. 2010. 'How to Evaluate Library Collections: A Case Study of Collection Mapping'. *Collection Building* 29 (2): 43–49. <https://doi.org/10.1108/01604951011040125>.
- IRUS-UK. 2018. 'International Collaboration and Value: Working with OAPEN'. IRUS-UK Case Studies. https://irus.jisc.ac.uk/documents/IRUS-UK_working_with_OAPEN.pdf.
- Kelly, Madeline. 2014. 'Applying the Tiers of Assessment: A Holistic and Systematic Approach to Assessing Library Collections'. *The Journal of Academic Librarianship* 40 (6): 585–91. <https://doi.org/10.1016/j.acalib.2014.10.002>.

- Lamothe, Alain. 2010. 'Electronic Book Usage Patterns as Observed at an Academic Library: Searches and Viewings'. *Partnership: The Canadian Journal of Library and Information Practice and Research* 5 (1): 1–16. <http://journal.lib.uoguelph.ca/index.php/perj/article/viewArticle/1071>.
- Lantzy, Tricia, Talitha Matlin, and Judy Opdahl. 2020. 'Creating a Library-Wide Collection Management Cycle: One Academic Library's Approach to Continuous Collection Assessment'. *Journal of Library Administration* 60 (2): 155–66. <https://doi.org/10.1080/01930826.2019.1677092>.
- Ogg, Lesley. 2010. 'Launch of the OAPEN Library at the Frankfurt Book Fair'. 20 September 2010. <https://blog.alpsp.org/2010/09/launch-of-oapen-library-at-frankfurt.html>.
- Perrin, Joy M., Le Yang, Shelley Barba, and Heidi Winkler. 2017. 'All That Glitters Isn't Gold: The Complexities of Use Statistics as an Assessment Tool for Digital Libraries'. *The Electronic Library* 35 (1): 185–97. <https://doi.org/10.1108/EL-09-2015-0179>.
- Snijder, Ronald. 2022. 'Big in Japan, Zimbabwe or Brazil – Global Reach and National Preferences for Open Access Books'. *Insights* 35 (June): 11. <https://doi.org/10.1629/uksg.580>.
- Stregger, Elizabeth, and Jaclyn McLean. 2021. 'Location-Based Trickery: Librarian Efforts to Provide Access to Unaffiliated Users'. *Weave: Journal of Library User Experience* 4 (2). <https://doi.org/10.3998/weaveux.222>.
- Whitford, Leslie. 2014. 'DOAB: Directory of Open Access Books'. *Reference Reviews* 28 (3): 12–13. <https://doi.org/10.1108/RR-10-2013-0269>.