



**Landscape study of small journal publishers for the Knowledge Exchange
Task & Finish Group for “Small Publishers and the Transition to Open
Access”**

German Centre for Higher Education Research and Science Studies (DZHW)

November 2022

This study was commissioned by Knowledge Exchange. The results and interpretations are the responsibility of the DZHW, as the institute conducting the study. However, the members of the Knowledge Exchange's Task and Finish group "Small Publishers and the Transition to Open Access", as listed below, were instrumental in designing the study and providing feedback on the results.

Frank Manista, KEG Lead, Jisc, United Kingdom

Jean-François Nominé, KEG Lead, CNRS, France

Saskia Woutersen, Expert group lead, Leiden University Library, The Netherlands

Margo Friederike Bargheer, University of Göttingen, Germany

Jesper Boserup Thestrup, Royal Danish Library - Aarhus, Denmark

Jean-François Lutz, University of Lorraine, France

Mafalda Marques, Jisc, United Kingdom

Emmanuelle Morlock, CNRS, France

Anna-Mette Morthorst, Royal Danish Library - Aarhus, Denmark

Antti-Jussi Nygård, The Federation of Finnish Learned Societies, Finland

Markus Putnings, University Library of Erlangen-Nürnberg, Germany

Jeroen Sondervan, Utrecht University, The Netherlands

Anne Thorst Melbye, University of Southern Denmark Library, Denmark

Timo Vilén, University of Helsinki, Finland

Ans ter Woerds, SURF, The Netherlands

Contact and further information:

German Centre for Higher Education Research and Science Studies (DZHW)

Dr Dimity Stephen

Schützenstr. 6a

10117 Berlin

Tel: +49 (0)30 2064177-26

Fax: +49 (0)30 2064177-99

Email: stephen@dzhw.eu

Contents

Introduction	1
Study Design	2
1. Classification of bibliometric data sources	2
Method	3
Results and conclusions	3
2. Identification of small publisher groups in KE countries	7
Method	8
Results	10
Conclusions and considerations	11
3. Open Access status of small publishers' journals	12
Methods	13
Results	14
Conclusions and considerations	24
4. Characteristics of small publishers	25
Results	26
Conclusions and considerations	33
5. Contrasting small publishers that have and have not transitioned to OA	33
Methods	33
Results	34
Conclusions and considerations	39
Overall conclusions	40

Introduction

The scholarly publishing system tends to be dominated by large corporate publishers that publish a majority share of academic research. Estimates from 2013 show that more than half of all papers in Web of Science (WoS) were produced by the top five publishers, and this increased to 70% in the social sciences [10]. The visibility and prominence of these publishers enables them to attract both individual researchers and research-performing institutions, for both of whom success is partly defined by the quantity of outputs published and the prestige of the publishing journal. However, small publishers play a vital role in scholarly communication. Small publishers offer essential channels of communication for, in particular, disciplines and academics with specific regional, cultural, linguistic, or thematic foci. This crucial bibliodiversity provided by small publishers facilitates richness in academic discourse by enabling participation by diverse actors with diverse ideas [16], fosters academic progress, and “develop[s] knowledge that would not exist at all if all publishing was to take place at the international level” [18].

The scholarly publishing system has begun to move toward more open accessibility of research material, led in recent years by the Plan S movement. Plan S aims to reduce the proportion of research obstructed by paywalls by encouraging research funders to require all research produced with their grants to be openly accessible. As cOAlition S, several major European, international and charitable research funders have committed themselves to Plan S, alongside a growing number of funders and universities [13]. However, small publishers face unique challenges in innovation and dissemination during this transition. Economics of scale penalises small publishers, as the necessary adaption of business processes to account for the Open Access (OA) transformation must be adopted to the few journals of small publishers, while larger publishers can not only reuse processes, but also learn from experience. At the same time, the OA transformation affects the whole publishing market, also opening new growth opportunities for small publishers [6].

The Task & Finish Group for “Small Publishers and the Transition to Open Access” of Knowledge Exchange (KE), advised and supported by the Knowledge Exchange office, commissioned this study to understand the landscape of small publishers in six European countries in the context of the transition to open science. This empirical examination of small publishers using bibliographic data sources and techniques enables the T&F group to make evidence-based recommendations to actors in the scholarly publishing system to sustain and support the bibliodiversity offered by small publishers during the transition to OA.

The German Centre for Higher Education Research and Science Studies (DZHW) has extensive experience in the development and application of journal analytics via its bibliometric research program and service activities. The DZHW usually applies the bibliometric database infrastructure of the German Competence Network for Bibliometrics (KB) as the basis for its research projects. The KB¹ is a project funded since 2008 by the German Federal Ministry of Education and Research (BMBF) that is dedicated to the establishment, expansion, operation, and maintenance of a quality-assured in-house version of the WoS and Scopus databases and the required analysis infrastructure for bibliometric applications. The KB managing office is operated by the DZHW. Based on the provision of this infrastructure, the KB enables the participating partners to carry out separate research and service activities. While these proprietary databases were not applied in this project due to the prioritisation of an open data approach, the experience gained with this infrastructure nevertheless informs the analytical design of the project.

¹<http://bibliometric.info/>

Study Design

This study presents an initial landscape on small publishers in the six KE member countries: Denmark, Finland, France, Germany, the Netherlands, and the United Kingdom. Small publishers are defined as publishers that produce up to 10 journals containing up to 240 articles per year. As with every map, a landscaping exercise requires a suitable coordinate system on which to locate the publishers. Hence, the study applies multiple, complementary data sources to observe the publishers' individual positions, where each data source serves as a unique coordinate axis highlighting a particular aspect of the overall landscape. Given their individual characteristics, publishers find themselves in different locations, while the overall spatial distribution of publishers on these coordinates supports the KE in identifying suitable candidates for further qualitative analyses and facilitates the development of recommendations for the support and continuation of small publishers during the transition to OA. To this end, the set of small publishers following an OA approach will be contrasted on the coordination system with the set of small publishers following a closed publishing model to identify the unique characteristics of small OA publishers and to highlight a potential way forward for publishers relying on closed, subscription journals.

Consequently, the study aims to answer the following questions:

1. How many small publishers in the six KE countries are identifiable in large bibliographic databases?
2. Given the individual characteristics of these small publishers, which spatial distribution arises on a coordination system of diverse, complementary bibliographical data sources?
3. How do small publishers that have completed a transition to OA differ on this coordination system from those that use closed publishing models?

The study consists of five Sections. The first Section finalises a literature study initiated by the T&F group as a background for the later analysis and reporting. To answer the first question, the second Section of the study examines the open database Crossref to identify small publishers, while additional data sources are applied to focus on publishers in the KE network countries. The third and fourth Sections answer the second question by providing additional information about the identified publishers, for example the OA status and a discipline classification of the respective journals. In parallel, the fifth Section contrasts the characteristics of small publishers that have and have not transitioned to OA to identify potentially key factors influencing this transition. Together the study's Sections establish the landscape of small publishers in Europe, facilitating the T&F group to develop evidence-based recommendations for publishers, policy-makers, and research-performing organisations to contribute to and support bibliodiversity and its sustainability.

1. Classification of bibliometric data sources

This Section completed desk research already begun by the T&F group to classify bibliometric data sources according to key features, and identify sources in which small publishers may be included. The basis of this investigation was 27 studies identified by the T&F group that examined small publishers. It was expected that potential data sources capturing small publishers could be drawn from these studies. Sources only rarely mentioned were of particular interest as they may highlight areas of academic literature that are typically un(der)examined and may be populated with small publishers. Identifying the data sources also sheds light on different perspectives used by researchers

to assess small publishers in the OA publishing landscape. As such, not only does this Section of the study highlight what type of data is utilised, but also helps us understand what the different sources offer, how they complement one another, and how this affects the understanding of OA publishing in light of the transformation that many small publishers are undergoing.

Method

We identified and coded all data sources, such as the Directory of Open Access Journals (DOAJ), mentioned in the 27 studies via a systematic reading of the studies, focusing on the characteristics of the research, such as the units of analysis, the subject-specific features of the studies, and the conceptual dimensions investigated. Specifically, we identified every data source in the 27 studies, and also how many different sources each study used, and the type of data included. We also looked at the features of each data source in relation to the specific objectives of each study, and described the units of analysis (articles, journals, or publishers). Given the different units of analysis, the datasets consequently vary in size, as smaller units (articles) produced larger datasets, while larger units (journals or publishers) produced smaller samples. Information about the studies' interconnection was examined by determining if the studies had cited or been cited by other studies in the sample. We then calculated the total number of data sources mentioned and the number of times each was mentioned in the studies and generated a matrix of the key features of the bibliometric data sources used. This can help differentiate between approaches, as well as highlight the practicality of the data used for each study regarding each research context or case study.

Results and conclusions

In total, the 27 studies utilised 36 data sources. These sources varied in their nature and characteristics. For example, many studies used primary data gathered through surveys and/or interviews, but most studies were based on analyses of secondary materials, such as databases (e.g. Scopus) or compiled datasets (e.g. the Diamond OA Journals study [2]). Further, most studies used more than one data source to substantiate their findings and, on average, three data sources were used in each analysis.

The diverse nature of the contexts studied is also relevant. Just under 40% of the studies examined the international context of OA journals and publishers, while around 15% each focused on more specific cases, namely, the Nordic countries, Germany or German-speaking countries, the Netherlands, or France. These characteristics of the studies provide context to the choice of data sources used. Around 60% of the studies examined OA publishing at the journal-level, whereas the remainder were divided into the publisher- or publication-levels. This distinction is relevant to understanding the differences in approaches to the analytical dimensions of the studies. It is also important to note that not all studies used bibliometrics to analyse OA publishing. Some were concerned with the institutional and structural characteristics of OA publishing in the context of a potential or ongoing transformative process and this may have influenced the data sources selected.

Supplementary Table S1 summarises the information pertaining to each study in this sample. Notably, this table includes the working definitions each study used when referring to small publishers. This information in particular highlights how diverse the concept of small publishers is: while some studies use revenue and costs in order to assess a publisher's size, others focused on output, measured as articles-per-issue-per-year, and several did not define small publishers beyond simply being "smaller than large or international publishers." Table 1 shows each of the data sources used, describes their use in each study and the level at which the data were examined.

Table 1: Overview of data sources used in sampled studies

Name	Use of data source in studies	Unit of analysis
Becerril et al. (2021)	OA Diamond journals study.	Journals
BFI	List of OA journals chosen by panels of academics. Includes low quality journals.	Journals
Bielefeld Gold OA	OA journals in Germany. May 2020 status.	Journals
Dimensions	Article-level OA information in OA and hybrid journals, and author funding information.	Articles/Journals
DOAJ	Data for OA journals based on size, license, archiving, copyrights and identifiers.	Journals
ESAC	Registry listings of transformative agreements (OA), 2020.	Journals/Publishers
Ferwerda et al. (2017)	Typologies of academic and non-academic publishers.	Publishers
German Publishers & Booksellers Assoc.	Database of German publishers. 2019 status.	Publishers
Gold OA (GOA)	OA Journals metadata and APC information.	Journals
Google Analytics 2014-2016	Access metrics for the journal.	Journals
Huber (2012)	Typologies of academic and non-academic publishers.	Publishers
Institutional data	Dutch public research institutional expenditure and licensing data. 2019 status.	Journals/Publishers
Interviews	Editorial management and business operations.	Journals/Publishers
JUFO	Finnish-produced rating/classification system of academic publication channels.	Journals
Literature review	OA Journals and Publishers' business operations.	Journals/Publishers
Morrison et al. (2019)	OA journals' APC information. 2019 status.	Journals
NOP-HS (2019-2022)	Grant funding information for journals 2019-2020.	Journals
NSD	List and ranks of Nordic countries' publications.	Journals
Open Access Directory (OAD)	List of OA-transformed journals.	Journals
Open Access Iceland	List of OA journals in Iceland.	Journals
OpenAPC	Dataset of fees paid for Open Access publishing.	Journals
Piwowar et al. (2019)	OA articles estimated growth (projected to 2025).	Articles/Journals
QOAM (van der Graaf, 2020)	Types of OA peer-reviewed journals, international, 2019.	Journals
ROAD	Data for OA journals based on size, license, archiving, copyrights and identifiers.	Journals
Scopus	OA peer-reviewed journals metadata and APC information.	Articles/Journals
Sherpa Romeo	List of publishing licenses.	Publishers
STM (Johnson et al., 2019)	Editorial management and business operations in OA journals.	Journals
Survey	Editorial management and business operations of OA journals.	Journals
Swedish National Library	Official list of journals and serial publications.	Journals
Tidsskrift.dk	List of published journals in Denmark.	Journals
Ulrichsweb	Active, academic, online-available, peer-reviewed journals and serial publications.	Journals
Unpaywall	Data for OA type, publication type, year and publisher information, matches DOIs to CrossRef.	Articles/Journals
VIRTA	Datahub compiling bibliographic information of publications produced at Finnish organisations.	Journals
Web searches	OA journal transformation data.	Journals
Wise & Estelle (2019)	OA transformative business models.	Journals/Publishers
WoS	OA peer-reviewed journals metadata and APC information.	Articles/Journals

To contextualise the data sources: BFI was the Bibliometric Research Indicator list of “the publication channels in which publication of research results are awarded points” in the Danish research system.² However, it was discontinued in December 2021.³ The Bielefeld Gold OA is a dataset produced by Bielefeld University of German Diamond OA journals.⁴ Bielefeld University Library also participates in OpenAPC, which “releases datasets on fees paid for [OA] journal articles by universities and research institutions under an open database license.”⁵ Dimensions, WoS, Scopus, Unpaywall, and Ulrichsweb are large, commercial or open bibliometric/bibliographic databases. DOAJ is “a community-curated online directory that indexes and provides access to high quality, open access, peer-reviewed journals.”⁶ ESAC is a collaboration coordinated by the Max Planck Digital Library in Germany to produce open resources.⁷

The German Publishers and Booksellers Association is “a trade association and a cultural organisation” that “promote[s] books and reading whilst also fostering fair copyright laws and the preservation of Germany’s policy of fixed book prices.”⁸ GOA refers to a dataset of Gold OA journals from the DOAJ.⁹ JUFO is the Publication Forum journal list, “a classification of publication channels created by the Finnish scientific community to support the quality assessment of academic research”.¹⁰ NOP-HS refers to the Joint Committee for Nordic research councils in the Humanities and Social Sciences, which is “a cooperation between research councils in Denmark, Finland, Iceland, Norway and Sweden responsible for research within the Humanities and Social Sciences”.¹¹

NSD refers to the Norwegian Register for Scientific Journals, Series, and Publishers, a list of approved scientific publication channels.¹² The Open Access Directory (OAD) is “a compendium of simple factual lists about OA to science and scholarship, maintained by the OA community at large”¹³ Open Access Iceland is a list of OA journals in Iceland.¹⁴ QOAM is the Quality Open Access Marker and “contains a comprehensive dataset on Open Access journals, aggregating data from JournalTOCs, DOAJ and CrossRef.”¹⁵ ROAD is the Directory of Open Access Scholarly Resources, a service of the ISSN International Centre to provide “free access to those ISSN bibliographic records which describe scholarly resources in Open Access sources”.¹⁶

Sherpa Romeo, operated by Jisc, is “an online resource that aggregates and analyses publisher open access policies from around the world and provides summaries of publisher copyright and open access archiving policies on a journal-by-journal basis.”¹⁷ STM is the International Association of Scientific, Technical and Medical Publishers, “the leading global trade association for academic and professional publishers.”¹⁸ Tidsskrift.dk is “the Royal Danish Library’s portal for the publication of professional, scientific and cultural journals in digital full-text.” Finally, VIRTAs refers to the VIRTAs Publication Information Service, which “contains information on research publications produced by Finnish

²<https://ufm.dk/en/research-and-innovation/statistics-and-analyses/bibliometric-research-indicator/bfi-lists>

³BFI: <https://medarbejdere.au.dk/en/pure/bfi>

⁴Bielefeld Gold OA: <https://pub.uni-bielefeld.de/record/2963331>

⁵OpenAPC: <https://treemaps.intact-project.org/page/about.html>

⁶DOAJ: <https://doaj.org/>

⁷ESAC: <https://esac-initiative.org/about/about-esac/>

⁸BdB: <https://www.boersenverein.de/english/>

⁹GOA: https://figshare.com/articles/dataset/Gold_Open_Access_2014-2019_GOA5_/12543080/1

¹⁰JUFO: <https://www.julkaisufoorumi.fi/en>

¹¹NOP-HS: <https://nos-hs.org/>

¹²NSD: <https://kanalregister.hkdir.no/publiseringsskanaler/Om>

¹³OAD: http://oad.simmons.edu/oadwiki/Main_Page

¹⁴OA Iceland: <http://openaccess.is/oaicelandic-journals/>

¹⁵QOAM: <https://www.qoam.eu/qoam-journal-composition>

¹⁶ROAD: <https://www.issn.org/services/online-services/road-the-directory-of-open-access-scholarly-resources/>

¹⁷Sherpa Romeo: <https://v2.sherpa.ac.uk/romeo/>

¹⁸STM: https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

universities, polytechnics and research organisations.”¹⁹

The number of times each data source was mentioned in the 27 studies is shown in Figure 1. Additionally, the data sources are classified according to the type of data used. The categories are a) *Database/hub*, if it corresponds to a portal hosting structured data, b) *Institutional*, if it refers to data gathered through organisational documentation, c) *Primary*, if the data were collected specifically for that study, either through surveys or interviews, and d) *Secondary*, if the data were collected from other existing studies.

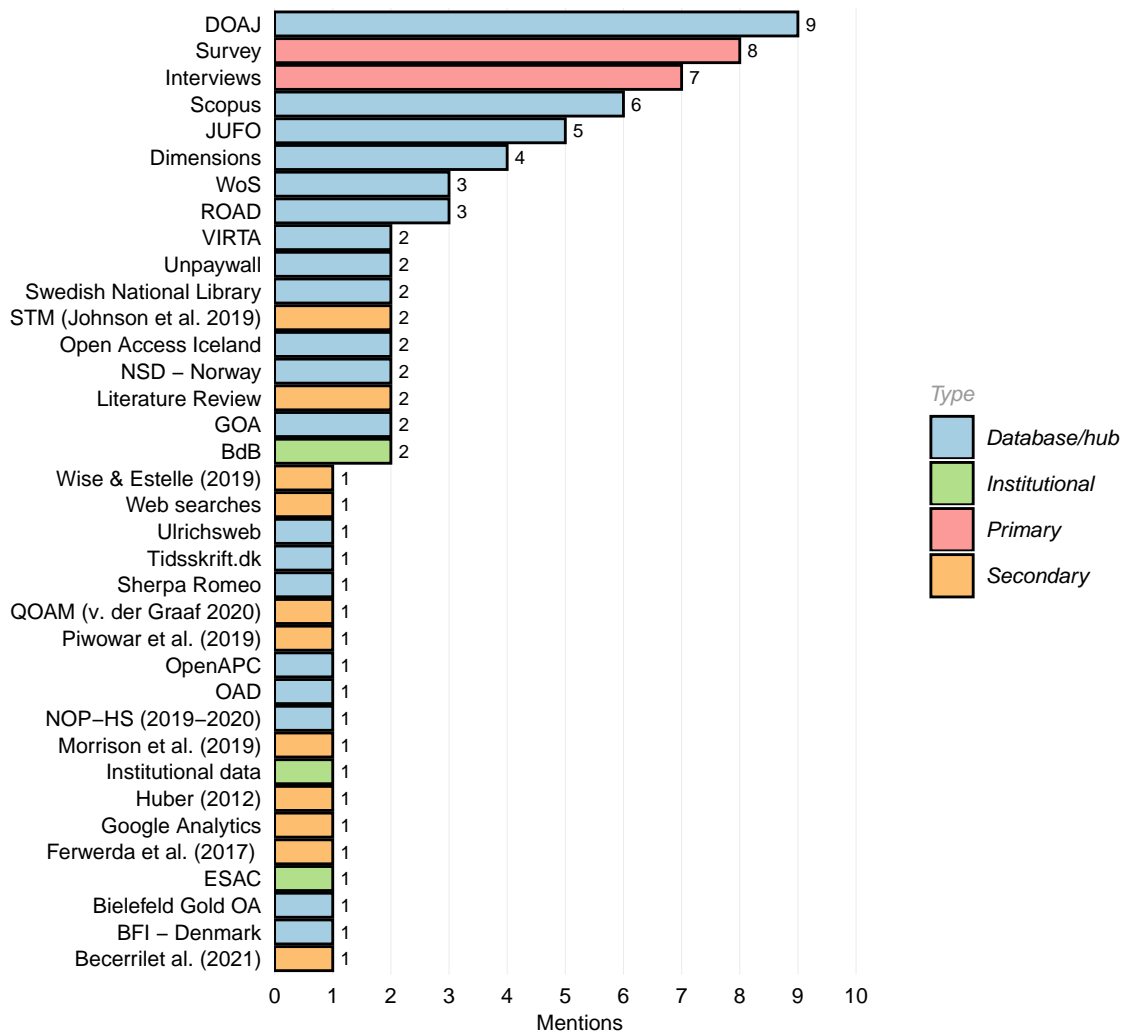


Figure 1: Number of mentions of each data source in the 27 studies.

Existing databases were the most common source for data about OA publishers and journals, and DOAJ in particular was the most frequently referenced data source. Surveys and interviews were also frequently used in the studies to directly obtain from journal editors and publishers information about their business logistics and editorial management. The large commercial databases were also often used, with Scopus mentioned twice as often as WoS, likely due to their different indexation policies leading Scopus to index a larger number of small publishers than WoS. It is noteworthy, however, that in most cases, data were contrasted with at least two other, different data sources to improve

¹⁹VIRTA: <https://www.csc.fi/en/-/virta-publication-information-service>

the quality of the final datasets analysed. In terms of identifying sources only rarely used and thus perhaps under-examined, several data sources were mentioned only once in the 27 studies. However, many of these, such as Tidsskrift.dk, NOP-HS, Bielefeld Gold OA, and BFI, have a national focus and thus may be of limited use for an examination of small publishers internationally. Conversely, sources such as QOAM and Ulrichsweb which are international in nature may be useful for further exploration of small publishers.

2. Identification of small publisher groups in KE countries

The aim of this Section of the study was to answer the first research question regarding how many small publishers in the KE countries are identifiable in a large bibliographic database, in this case Crossref. The publishers identified constitute the dataset upon which the remainder of the analyses were performed. Crossref was selected as the basis for this task for two reasons. First, Crossref membership terms require that its members only provide content likely to be cited in the scholarly ecosystem and hence separates scholarly from non-scholarly content. Second, Crossref has become the predominant DOI agency for publishers in the KE membership countries. Hence it constitutes the largest and most comprehensive collection of bibliographic material for KE membership countries. Although not all publishers will be registered with Crossref, the use of Crossref DOIs is becoming a common practice for publishers: Crossref has recorded continuous growth in membership since its establishment, with the majority of its members joining in the lower fee tiers, indicating they are likely small publishers (p. 7, Crossref [5]). As such, Crossref provides a strong, publicly available foundation for identifying small publishers in the KE countries.

The classification of publishers as small and attributable to a KE country is complex for several reasons. First, the number of journals has previously been used to categorise publishers by size. However, there is no standardised threshold for what constitutes a small publisher. For example, Laakso and Multas [9] defined small- and mid-sized publishers as those that produced up to 150 journals, while Pacher [12] classified publishers that produced more than 30 as “major”. It was decided with the T&F group to apply the following classification to identify and differentiate groups of small publishers, which is similar to the classification used by the OA Diamond Journals study [3]:

1. Very small publishers: produced 1 journal per year and not more than 102 articles per year
2. Small publishers: produced 2-5 journals per year and not more than 240 articles per year across all journals
3. Rather small publishers: produced 6-10 journals per year and not more than 240 articles per year across all journals

The document-level thresholds were decided in collaboration with the T&F group and were derived from the mean and median number of articles produced per year by journals in WoS and in Crossref. Chen [4] reported that journals indexed in WoS’ Science Citation Index (SCI), i.e. “hard” science disciplines, published a median of 102 and mean of 175 articles per year, while journals in the Social Science Citation Index (SSCI) published a median of 32 and mean of 82 articles per year. Journals in Crossref, which indexes smaller publishers often not captured in commercial databases, published a median of 24 and mean of 66 articles per year. As such, 240 articles per year was decided as the threshold for document counts on the basis of the median articles of journals in Crossref multiplied by 10 as rather small publishers may publish up to 10 journals. Further, the threshold of 102 documents per annum for very small publishers was decided based on the median annual articles of journals in

the SCI. Throughout this document, “small publishers” refers to all three categories of publishers, unless otherwise specified.

A second complexity in identifying relevant publishers is that the concept of national publishers is empirically fuzzy and can be operationalised in many ways [7]. This may include, for instance, the publisher’s registered location, the location of authors who publish in the publisher’s journals, the location of authors citing published material, or the indexation of the publisher in library catalogues, all of which signal relevance of the publisher and its journals to particular countries. As authors’ affiliation data are known to be extensively missing in Crossref, we relied instead on the ISSN registrant’s country in the ISSN Portal²⁰ to assign a journal’s country. However, a limitation of this method is that this location may be potentially defined by business necessities of the publisher, rather than its local relevance to the academic community.

Method

Crossref data were obtained via a data file provided in May 2022 and consisted of 134 million records registered until the end of April 2022. These data were converted to PostgreSQL data tables stored on the DZHW’s infrastructure. We then extracted from the Crossref data the details of all publishers with journals produced in 2019-2021, as the latest three complete years of data. Publishers were disambiguated on the basis of the Crossref member identifier. This identifier more accurately disambiguated publishers than did DOI prefixes or publisher names, as Crossref members might hold several prefixes and publisher names do not account for linkages between businesses. Conversely, the membership identifier was consistent and unique to each publisher and accounted for commercial connections. Journals were disambiguated based on their titles, as opposed to ISSNs, as journals usually held more than one ISSN for their print and online versions. Extensive manual cleaning of the titles was undertaken to remove instances of multiple titles per journal due to spelling variations, e.g. “Journal of food chemistry *amp* nanotechnology” and “Journal of food chemistry *and* nanotechnology”, or the inclusion of volume or issue details in the title. This cleaning reduced the dataset to one title entry per journal.

Based on the cleaned title data, publishers were then classified on the basis of the number of journals they published according to the classification outlined above. Further, the number of documents produced by publishers across all journals were then examined to identify and exclude publishers that produced a small number of high-output journals, e.g. PLoS, and consequently would not be considered a small publisher. It must be acknowledged that, as all items included in a journal are classified in Crossref as journal articles, these document counts included a small number of editorials and other non-research material.

To supplement possible gaps in the Crossref data, KE group members provided lists of journals and publishers pertinent to researchers in the KE countries. Three lists in relation to journals used by researchers in the Netherlands were provided. The first consisted of 131 unique publishers with journals published since 2018 based on entries in the Dutch UKBsis database²¹, a scholarly information system containing academic content from the Dutch university libraries. The journals were matched to the Crossref dataset by ISSN to identify new journals/publishers not in Crossref and then internet-based research was conducted to identify if these publishers were considered small based on the number of journals they published. Journals from 41 additional publishers from this list were incorporated

²⁰The ISSN Portal, a partially-open source database operated on a not-for-profit basis, contains metadata of registered ISSN entities.

²¹<https://ukb.nl/english/ambitions/>

into the Crossref-based dataset. Two further lists based on journals published in by researchers at the Universities of Leiden and Amsterdam in 2019-2021 were also supplied. The matching and internet research process was repeated and identified a further 84 small publishers. Notably, the number of documents produced by these journals annually was unknown – as this information could not be sourced from Crossref as we had for the other journals – and as such it is possible some publishers may be out of scope based on document-level output.

We also received a list of 1,121 journals from 644 publishers relevant to French academics extracted from the Cairn²² and OpenEdition²³ platforms. These platforms are collections of humanities and social science publications and Cairn has a specific focus on francophone content. We matched the ISSNs from this list to the Crossref data to identify additional journals/publishers and conducted internet-based research to classify the publishers. This identified 29 additional small publishers. However, this list was particularly beneficial as it clarified the publisher of a substantial number of journals for which the publisher was recorded as Cairn or OpenEdition in Crossref. For these publishers identifiable in Crossref, document counts from Crossref were examined and 18 such publishers were excluded as they exceeded the document threshold. However, we could not obtain document counts for introduced publishers that were not in Crossref. As such, some publishers that exceeded the document-level output threshold may have been included. After inclusion of the French and Dutch lists, the dataset consisted of 23,058 journals published by 14,328 publishers, of which 10,772 were very small publishers, 3,044 were small publishers, and 512 were rather small publishers. For context, in total there are 19,434 publishers in Crossref, based on member IDs.

Due to resource constraints, it was necessary to reduce the number of journals that were to be examined via the ISSN Portal in the next step. It was examined whether, using Crossref data, journals could be excluded on the basis the majority of articles were written in non-European languages and/or the publisher was located outside of Europe. Ultimately, however, these data were too incomplete to facilitate this filtering. As an alternative, the ISSNs of small publishers' journals were matched to the DOAJ to obtain the publishers' countries. A publisher was attributed the country recorded in DOAJ on the basis that at least one journal matched by ISSN. From this process, 3,300 journals were excluded as their publishers were not located in Europe, the United States of America, or Canada. The United Nations definition of Europe²⁴ was used to define the region. Publishers in the USA were retained as the USA's prominence as an academic system may mean USA-based journals are relevant to European researchers. Similarly, Canadian publishers were retained so as not to exclude publishers that are hosted in Canada but particularly relevant to French researchers due to the shared language. The dataset was thus reduced to 19,574 journals published by 12,622 publishers, of which 9,680 were very small publishers, 2,558 were small publishers, and 384 were rather small publishers.

Finally, to further refine the group to small publishers in member countries of the KE, we called the ISSN Portal API via R for one ISSN per journal to identify the publishers' registered countries. There were 933 journals that did not have ISSNs and so we manually searched these journals' titles via the ISSN Portal website and retrieved the publishers' countries. Thirty-one journals were published in one of the KE countries and we retrieved the complete ISSN Portal data for these additional records via the API. The following variables were obtained from the ISSN Portal for use in this Section of the study and the later Section analysing other characteristics of the relevant small publishers: start and end years, accrual periodicity, languages accepted, subject classifications, publisher, publishers country, and the publisher of the journals' earliest publication events.

²²<https://www.cairn-int.info/about.php>

²³<https://www.openedition.org/10918>

²⁴<https://unstats.un.org/unsd/methodology/m49/>

A publisher was considered relevant to a KE country when any of its journals were published in that country. Publishers could be allocated to more than one country when journals were published in two or more KE countries. However, it must be noted that the names of the publishers recorded between Crossref and the ISSN Portal were not always consistent. We compared the publisher names between data sources for the journals identified as relevant to the KE countries and found that 66% of the publishers aligned. However, a third of journals had different publishers recorded in each source. Due to resource constraints and the relatively large sample size involved, it was infeasible to manually resolve discrepancies to determine which source was correct. Consequently, for consistency as Crossref is the basis of the study, we accepted the publisher as recorded in Crossref and the country of the journal as recorded in the ISSN Portal. This means that there may be some error in the inclusion of publishers as relevant to a country. However, inconsistent publisher names between data sources do not necessarily indicate that country allocations were incorrect. For instance, one journal was recorded as published by the *Universitätsverlag Göttingen* in one source and by the *Zentrum für USA Studien* in the other. Another was attributed to The Finnish Society of Photogrammetry and Remote Sensing or the Helsinki University of Technology. The allocation of the journals to Germany and Finland respectively is thus likely accurate, despite the different publisher names between sources. To provide an external validation of the accuracy of the allocation of publishers to countries, members of the T&F group manually examined the journals assigned to each country (except the United Kingdom, unless the journal was also assigned to a second country) and indicated whether they believed the journal was in/correctly assigned or they could not be certain.

Results

After refining the sample to publishers affiliated with the six KE countries, we identified 1,029 small publishers of 1,696 journals. This was comprised of 45 rather small publishers publishing 322 journals, 195 small publishers publishing 585 journals, and 789 very small publishers of 1 journal each. The number of publishers by category and country are shown in Table 2. Please note that, as some publishers are assigned to more than one country, summing counts of journals or publishers across countries does not align with the totals provided. Publishers are always assigned to only one size category, however. A complete list of the publishers, their journals, their size classification, and affiliated countries is available in Supplementary Table S2 and Table S3 provides an overview of the number of publishers and journals by both publisher category and country.

Table 2: Counts of publishers and journals by category and country

Country	Rather small	Small	Very small	Total publishers	No. journals
Denmark	1	2	12	15	25
Finland	0	7	86	93	106
France	17	59	393	469	687
Germany	8	33	73	114	236
Netherlands	2	26	61	89	147
United Kingdom	18	75	168	261	527
Total	45	195	789	1029	1696

Interestingly, publishers from the KE countries comprised a relatively small proportion of the small

publishers identified in Crossref. Of the over 19,000 journals identified, 8.6% were attributable to publishers in the KE countries. Conversely, Indonesia (18.1%), South Korea (9.6%), the United States (7.4%), Russia (6.9%), Turkey (6.0%), and Japan (4.1%) accounted for large proportions of journals. This may reflect particularities of the publishing systems in these countries or their likelihood to use Crossref to register DOIs and hence affects their visibility in centralised infrastructures like Crossref and subsequent products such as the bibliometric database Dimensions. For instance, anecdotal evidence suggests German publishers often use alternative DOI registries to Crossref, and thus may not be captured in this sample. Further, that the highest number of publishers were identified in France is partially attributable to the supplementary list provided encompassing journals in OpenEdition and Cairn. This list was instrumental in disambiguating many small French publishers that were attributed to OpenEdition or Cairn in Crossref.

The results of the T&F group members' validation are shown in Table 3. These data indicate that, for instance, 69% of journals allocated to Germany based on data from the ISSN Portal were correctly assigned, 22% were incorrectly assigned to Germany, and the members were uncertain about 9% of journals assigned to Germany. For each country, the majority of journals were judged to be correctly assigned to that country. However, nearly a quarter of journals attributed to Germany and the Netherlands were believed to be inaccurately assigned. These results should be considered when interpreting the remaining results in this report. Further, this validation offers no information about the journals not captured in the identification process. For instance, T&F group members from Denmark believe the number of Danish publishers and journals is higher than presented here. This is believed to stem from a similar circumstance as observed with OpenEdition and Cairn. Here, the Royal Danish Library is recorded in Crossref as the publisher of a large number of journals, some of which are actually produced by small publishers and thus they were unable to be captured in this study.

Table 3: The percentage of journals in/correctly assigned to each country as assessed by T&F group members

Country	Correct	Incorrect	Unsure	Not checked
Denmark	100.0	0.0	0.0	0.0
Finland	91.5	8.5	0.0	0.0
France	91.4	5.5	0.6	2.5
Germany	68.6	22.0	9.3	0.0
Netherlands	70.1	25.2	4.8	0.0
United Kingdom	1.5	0.0	0.0	98.5
Total	59.2	7.9	1.9	31.0

Conclusions and considerations

Applying Crossref as the principal data source brings, due to its characteristics and practices among publishers, the risk of potential false positives and false negatives. For instance, publishers holding several prefixes or, from a business owner perspective, several memberships might result in falsely declared small publishers. Erroneous or incomplete registration practices by publishers also hinder their identification, as does the loose Crossref membership policy. Further, although Crossref has demonstrated its position as the predominant DOI agency for KE countries, alternate agencies exist and the study's prerequisite of a Crossref DOI likely excludes some small publishers in the KE

countries that use other DOI agencies, e.g. DataCite, to identify their content or do not register for DOIs at all. Unfortunately, the extent of small publishers missing from Crossref cannot be quantified or accounted for within the scope of this project. However, the inclusion of French-language journals and journals used by Dutch authors helped to identify 86 small publishers (8.4% of all publishers in the sample of 1,029) missing from Crossref. It is thus recommended that similar sources of language- or country-specific data sources are implemented in future studies to increase the coverage of small publishers.

As outlined above, the assignment of a publisher to a particular KE country is challenging for several reasons. We have applied here the method of identifying national journals by the country in which their publisher is registered in the ISSN Portal and this appears to largely be accurate. However, we acknowledge that this represents a specific approach to this task and future studies may choose to expand this approach to consider other methods through which nationality can be attributed, such as the authors publishing in or citing the journals' content.

Further, due to the large sample size of the study, it is unfeasible to reconcile variations between data sources. As such, we adopt here the lens of Crossref when viewing and interpreting the data in this study. Due to differences in sources from which databases draw their information, standardisation and other practices, each data source presents its unique lens through which we can interpret the landscape of small publishers and it cannot readily be said whether one lens is more accurate than another. As such, we remind readers that this report presents data from this Crossref lens, supplemented by additional sources, and this influences the results and conclusions drawn here.

3. Open Access status of small publishers' journals

The aim of this Section of the study was to explore the OA characteristics of the publishers and their journals according to several complementary bibliographic data sources. Here, we examine what information about the OA practises of the publishers can be derived from Crossref and Sherpa Romeo, and examine the information pertaining to OA statuses in DOAJ and Unpaywall.

The DOAJ “is a community-curated online directory that indexes and provides access to high quality, open access, peer-reviewed journals”.²⁵ It is an independent body, financially supported by a diverse assortment of libraries, publishers, and organisations, whose mission is to “increase the visibility, accessibility, reputation, usage and impact of quality, peer-reviewed, open access scholarly research journals globally, regardless of discipline, geography or language”. As of June 2022, the DOAJ indexed 17,745 journals spanning all disciplines, countries, and languages. A publisher may apply for indexation in the DOAJ providing it fulfils the following inclusion criteria: i) its journal(s) is OA, ii) it has published 5 or more scholarly research articles per year for at least one year, iii) and the journal(s) has a website, an ISSN, an editor and editorial board, and a peer-review process. OA is defined by the DOAJ as the publishing of the full text of an article for free and without delay using a license that allows users to access and use the content for any lawful purpose. The presence of a journal in the DOAJ certifies that the journal publishes its content OA.

As previously described, Jisc's Sherpa Romeo is “an online resource that aggregates and analyses publisher open access policies from around the world and provides summaries of publisher copyright and open access archiving policies on a journal-by-journal basis.”²⁶ Sherpa Romeo is one component of a larger suite of services that aim to provide guidance to researchers and institutions about OA

²⁵<https://doaj.org/>

²⁶<https://v2.sherpa.ac.uk/romeo/>

publication and compliance practises of journals and publishers. Publishers typically must apply to Jisc for inclusion in Sherpa Romeo. However, Jisc may strategically select publishers if they believe their inclusion would be beneficial (personal communication, 27 June 2022). As of June 2022, Sherpa Romeo held data for 4,263 publishers. The majority of these were located in the United States (889) and the United Kingdom (530), however there were also 136 German publishers, 83 French publishers, 53 Finnish publishers, and 48 Dutch publishers present in the database.

Unpaywall is a service of OurResearch, a non-profit organisation with the mission to improve the accessibility of scholarly research. Unpaywall amasses content with DOIs from more than 50,000 sources, including Crossref, DOAJ, PubMed Central, DataCite, preprint repositories and journals. It seeks to assist researchers in finding OA or legally archived versions of articles. Unpaywall also collects the OA status of its over 32.3 million articles, providing a useful dataset for the analysis of the OA status of publishers' content. Using DOAJ, Sherpa Romeo, and Unpaywall, we can assess the coverage of small publishers in these sources and develop a comprehensive picture of the OA characteristics of small publishers in the KE countries.

Methods

In examining the metadata of Crossref, we identified that the information pertaining to OA status in Crossref consisted of a variable directing users to a URL for the license of the article used in publishing, e.g. <http://creativecommons.org/licenses/by/3.0>. As such, we examined OA status at the article level. We first extracted from Crossref the DOIs and license information of all articles published between 2019 and 2021 in the journals of small publishers identified in Section 2 of the study. Data were extracted on the basis of Crossref member identifiers for publishers within Crossref, or via the ISSNs of publishers reassigned from Cairn/OpenEdition to more specific publishers on the basis of the French supplementary list. However, 111 journals introduced via the supplementary lists were not present in Crossref. As such, no DOIs could be obtained for these journals and they were excluded from this analysis. A further 7 journals did not have DOIs available in Crossref, although they were Crossref members, as their content was classified as other material, and these were also excluded.

After extracting all DOIs for the small publishers' documents, we sought to exclude from the sample front and back matter, such as introductions, editorials, book reviews, instructions for authors/reviewers, etc, as this material may have different OA statuses than research material. We excluded these documents by searching the titles for keywords such as editorial, instructions for authors, etc, and removed items containing them.

We then interpreted the license information to assign the document an open or closed access status. In a previous study[1], the DZHW examined the licenses referenced in Crossref and attributed them to an OA status of open or closed based on an assessment of the licenses' associated conditions. We utilised this classification in the present study to attribute each document an OA status. A small number of available licenses were not included in the previous study, and so the OA status of documents with these licenses could not be determined within the scope of this study.

To assess the OA information available in other sources, we then matched our sample of small publishers to the DOAJ, Sherpa Romeo, and Unpaywall datasets. We obtained journal-level data from the DOAJ from the public data file²⁷ for 7 June 2022. We then matched journals from the DOAJ to small publishers' journals based on any ISSN listed for the journal in Crossref. Matching via ISSN

²⁷<https://doaj.org/docs/public-data-dump/>

might be imperfect due to errors in the recorded ISSN in either source and this might have had a negative effect on recall, but only a limited effect on the precision of the analysis. For journals that did not have an ISSN, we fuzzy matched the journal titles between the small publishers and DOAJ datasets using the `fuzzyjoin` package [15] in R [14] and a threshold Levenshtein distance of 6 [11]. We manually confirmed matches on the basis of the title match and also similarity in the publishers' names. We determined from matching to the DOAJ an indicator of whether the journal/publisher published OA on the basis of its presence in the DOAJ, and also the year in which the journal started publishing OA.

To extract data from Sherpa Romeo, we used the Sherpa Romeo API to search the database for all ISSNs listed for the small publishers' journals in Crossref. To include the journals without ISSNs, we extracted all publishers in Sherpa Romeo from the website²⁸ and then fuzzy matched these with the publisher names of those journals missing ISSNs. We used a threshold Levenshtein distance of 4, in contrast to 6 for matching journal titles with the DOAJ, as publisher names were shorter in length than journal titles. Once again, the matches were manually confirmed and the relevant data for the additional journals identified through this process were extracted manually from the Sherpa Romeo website.

The data obtained from Sherpa Romeo consisted of the copyright owner, requirement for additional OA fees, conditions, embargo period, and license information for each of the submitted, accepted, and published article versions for each journal, in addition to an indicator of the presence of the journal in the Sherpa Romeo database to assess coverage.

To access Unpaywall data we used the `roadoi` package [8] in R [14] to call the Unpaywall API. DOIs are the basis of calls to this API so we used the sample of DOIs identified from Crossref to retrieve the OA status of each document (closed, gold, bronze, green, hybrid) and derived an indicator of the presence of the journal in the Unpaywall database. As noted, this sample excluded 118 journals without DOI or articles in Crossref. From these data, we calculated the percentage of each journal's documents that were of each OA status, and then allocated an overall OA status of open, closed or mixed for the journal; the journal was allocated as open or closed when 100% of the documents were open or closed, otherwise journals were considered mixed. This indicator distinguishes mixed journals from open and closed journals, in comparison to Unpaywall's journal-level OA indicator, which cannot distinguish between mixed and closed journals. Similarly, publishers were determined to be open, closed, or mixed on the basis of the classification of their journals. However, because OA status was not examined annually, journals that changed OA status during the 2019-2021 reference period – for instance from mixed to completely OA – were not identifiable and would be classified as mixed.

Results

The top row of Figure 2 shows the number and percentage of documents by OA status and publisher category, based on the documents' licenses in Crossref. Similarly, the bottom rows show the number and percentage of documents by OA status and publisher country. We see that the majority of documents (82.7%) did not have license information available. The aforementioned previous study by the DZHW examining licenses in Crossref similarly noted that nearly half of documents did not have license information[1]. The greater percentage of documents missing this information in this sample may reflect the focus on small publishers. This pattern persisted across countries, with approximately

²⁸https://v2.sherpa.ac.uk/view/publisher_list/1.html

75% of documents from each country missing license information. However, license information from Finnish publishers was somewhat more complete (58.7% missing), and almost always missing for French publishers (98.1% missing). Due to the incompleteness of these data, only a limited interpretation of this information can be made.

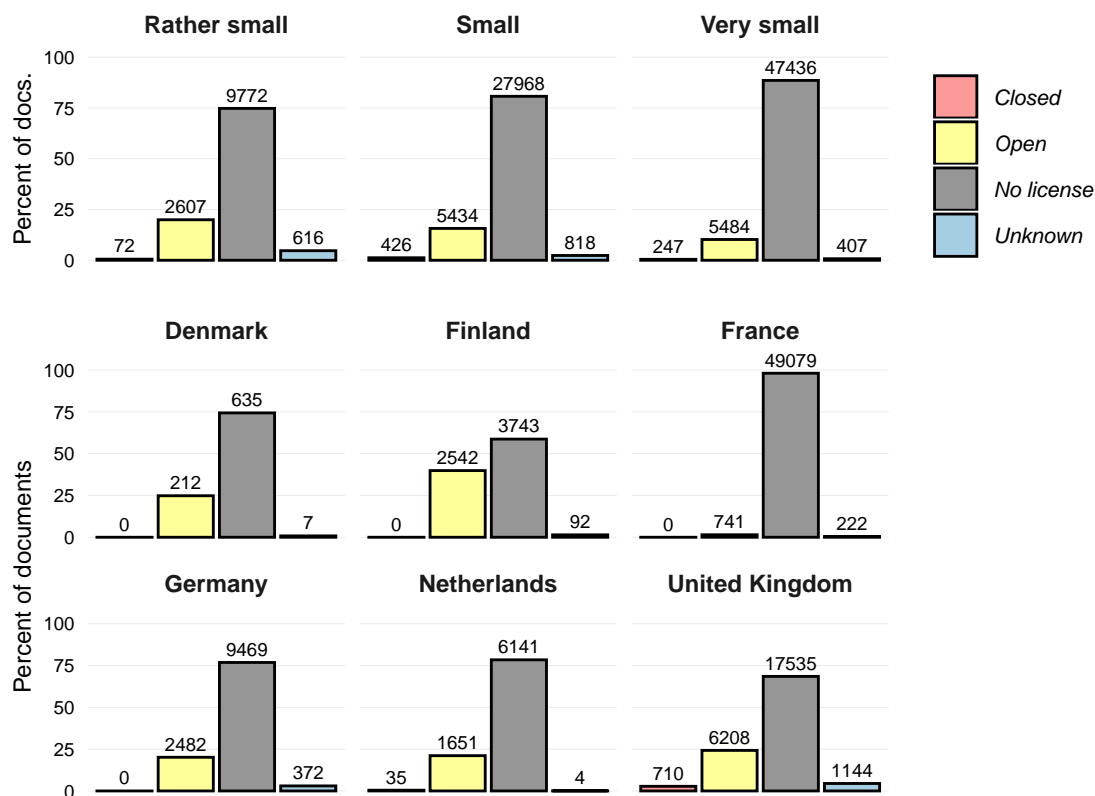


Figure 2: Number and percentage of documents by Crossref OA status, publisher category, and country.

Moving then to the complementary sources of OA information, we first examine the coverage of the sample within these sources. Of the 1,696 journals in the sample, 404 journals (23.8%) from 322 publishers (31.3%) were present in the DOAJ, 358 journals (21.1%) from 234 publishers (22.7%) were present in Sherpa Romeo, and data for 1,577 (93.0%) journals from 969 publishers (94.2%) could be retrieved from Unpaywall. As 118 journals could not be searched in Unpaywall due to a lack of availability of DOIs, coverage is actually closer to complete as only 1 journal was not present in Unpaywall. The essentially complete coverage of the sample in Unpaywall reflects the interlinkages between the datasets, with Crossref used as a source of information for Unpaywall. Table S2 in the Supplementary Tables contains indicators of the presence of each journal in the DOAJ, Sherpa Romeo, and Unpaywall, as well the number and percentage of documents by each OA status, and the overall status assigned to the journal and publisher.

The number and percentage of publishers present in the DOAJ by category is shown in the top row of Figure 3 and by category and country in the bottom two rows. Publishers were considered present if at least one of their journals was indexed in the DOAJ. The presence of a publisher in the DOAJ indicates it publishes one or more journals as OA. Twenty of the 45 rather small publishers (44.4%)

had journals indexed in the DOAJ and were better represented than small (37.9%) or very small publishers (28.9%). However, rather small and small publishers were more likely to be classified as present in the DOAJ or Sherpa Romeo as inclusion depended on only one of their 2-10 journals being present compared to the very small publishers' single journal. As such, we also show the journal-oriented perspective of coverage in the DOAJ in Figure 4.

This pattern of somewhat more rather small and small publishers covered than very small publishers was also present for most countries. The percentages of each country's publishers covered were also similar within categories. Variations in the percentage of content in the DOAJ between publishers of different categories and countries might reflect the practices of publishers or the importance attributed to applying for inclusion in the DOAJ. For instance, a recent project of the Federation of Finnish Learned Societies to encourage OA journals to become indexed in the DOAJ may have increased the coverage of Finnish publishers and journals.²⁹

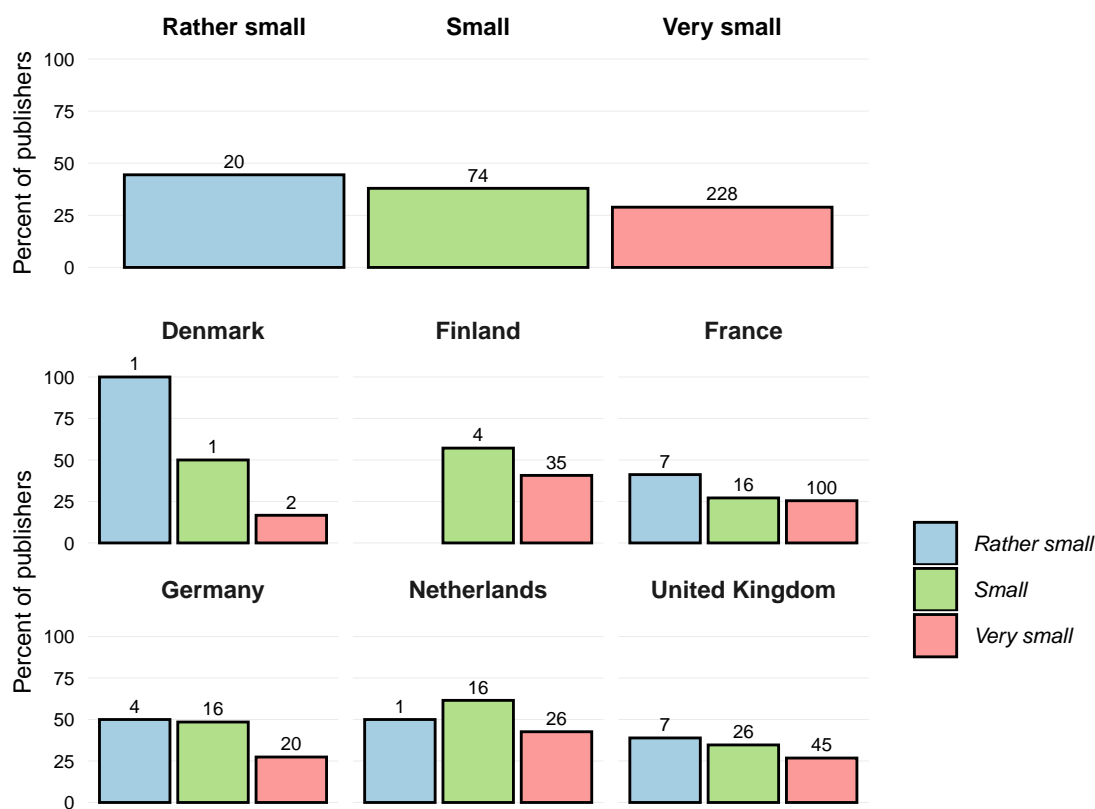


Figure 3: Number and percentage of publishers in DOAJ by publisher category, and country.

In contrast to this publisher-oriented perspective, we present in Figure 4 the number and percentage of *journals* that were in the DOAJ by publishers' country and category. Here, very small publishers' journals were somewhat better represented (28.9%) than journals from small (19.7%) and rather small publishers (18.9%). Also, in general, larger percentages (35-50%) of Finnish, Dutch, and Danish publishers' content were in the DOAJ than French, German and British (25% or less). Overall, this perspective highlights that many small and rather small publishers may have one or more journals in the DOAJ, but typically do not publish their whole corpus as OA.

²⁹<https://blog.doaj.org/2019/09/02/new-pilot-to-encourage-finnish-open-access-journals-to-apply-to-doaj/>

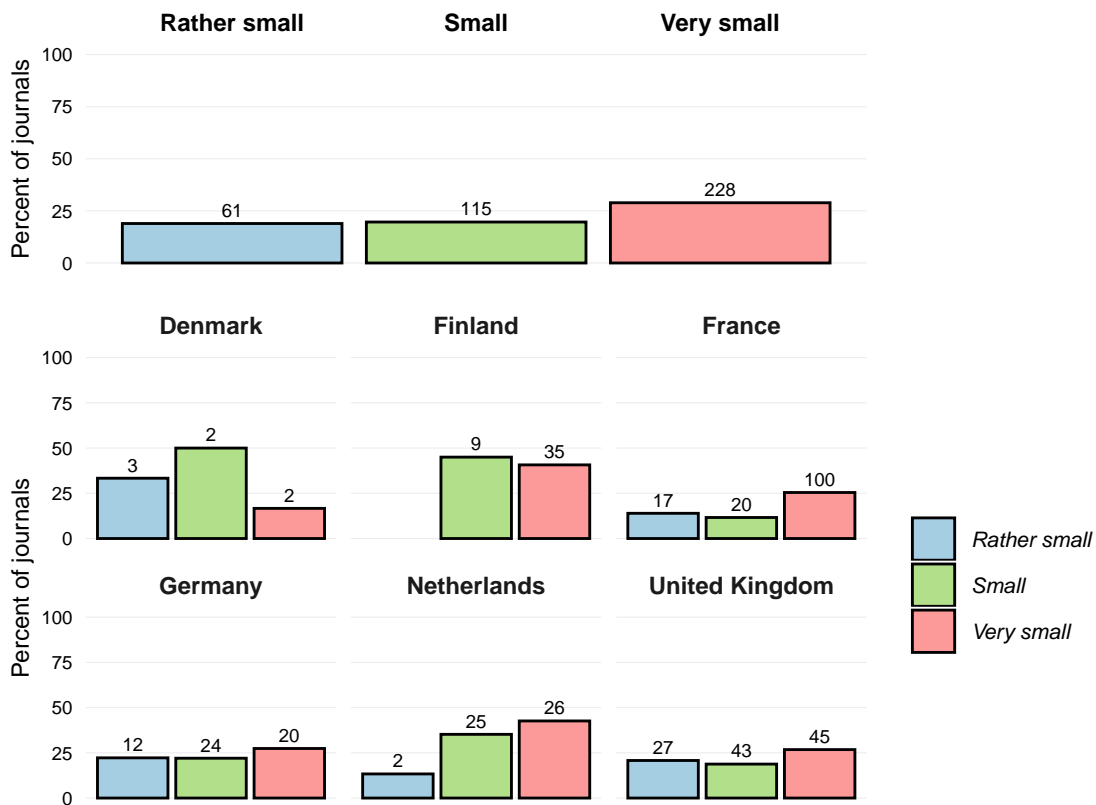


Figure 4: Number and percentage of journals in DOAJ by publisher category, and country.

Figure 5 shows the number and percentage of publishers that were identified in Sherpa Romeo by country and publisher category. Being present in Sherpa Romeo does not reflect a particular OA status. As with the DOAJ, publishers were considered present in Sherpa Romeo when at least one of their journals was included and so small and rather small publishers had an advantage here. As such, very small publishers had the lowest representation in Sherpa Romeo with 18.5% present, compared to a third of rather small publishers and 37.4% of small publishers. German and in particular French publishers were least represented, with typically a quarter or less of publishers of each category present, although this increased to 42% of small German publishers. Half or more of rather small and small Dutch, British, Finnish, and Danish publishers had at least one journal in Sherpa Romeo, and approximately a quarter of these countries' very small publishers were present. However, overall, there was no obvious skew toward any of the six countries' publishers.

In contrast to the publisher-oriented perspective, Figure 6 shows the number and percentage of *journals* in Sherpa Romeo by publisher category and country. The publisher- and journal-oriented perspectives of inclusion in Sherpa Romeo are quite similar, although the percentage of journals covered is somewhat lower than the percentage of publishers. As such, the publisher-oriented perspective somewhat inflates the perceived level of coverage as not all of the publishers' journals are present. The coverage of German and particularly French journals was notably lower than the other countries, but again the small sample of Danish journals should be considered.

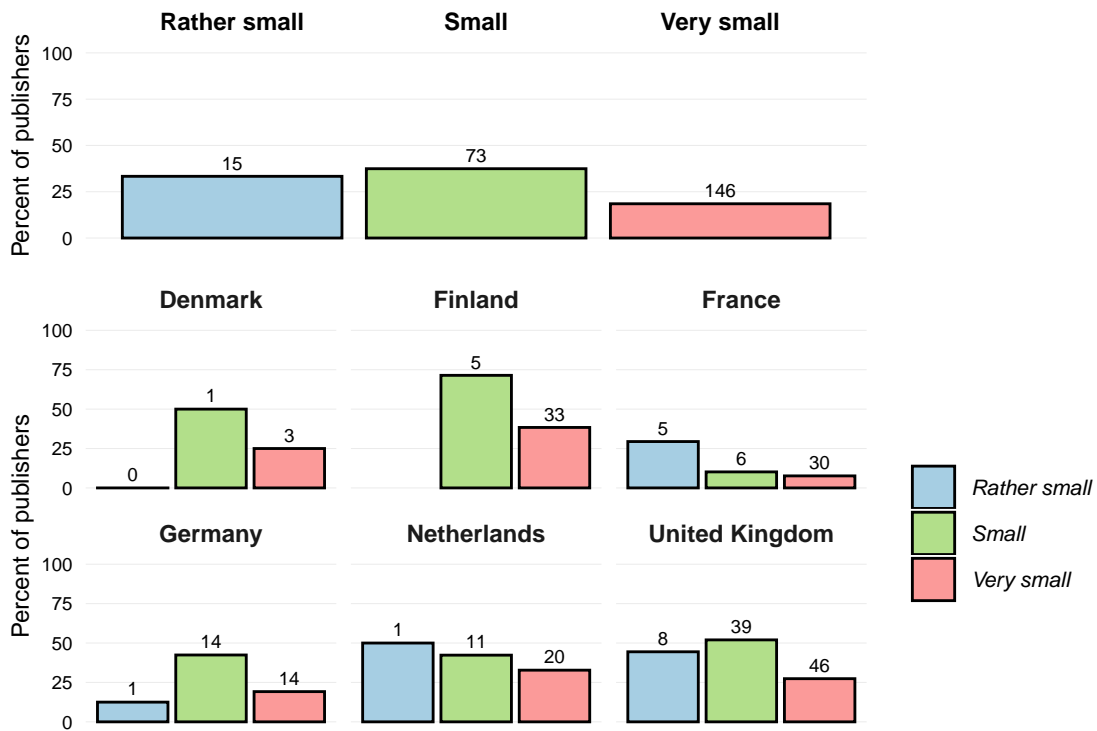


Figure 5: Number and percentage of publishers in Sherpa Romeo by publisher category, and country.

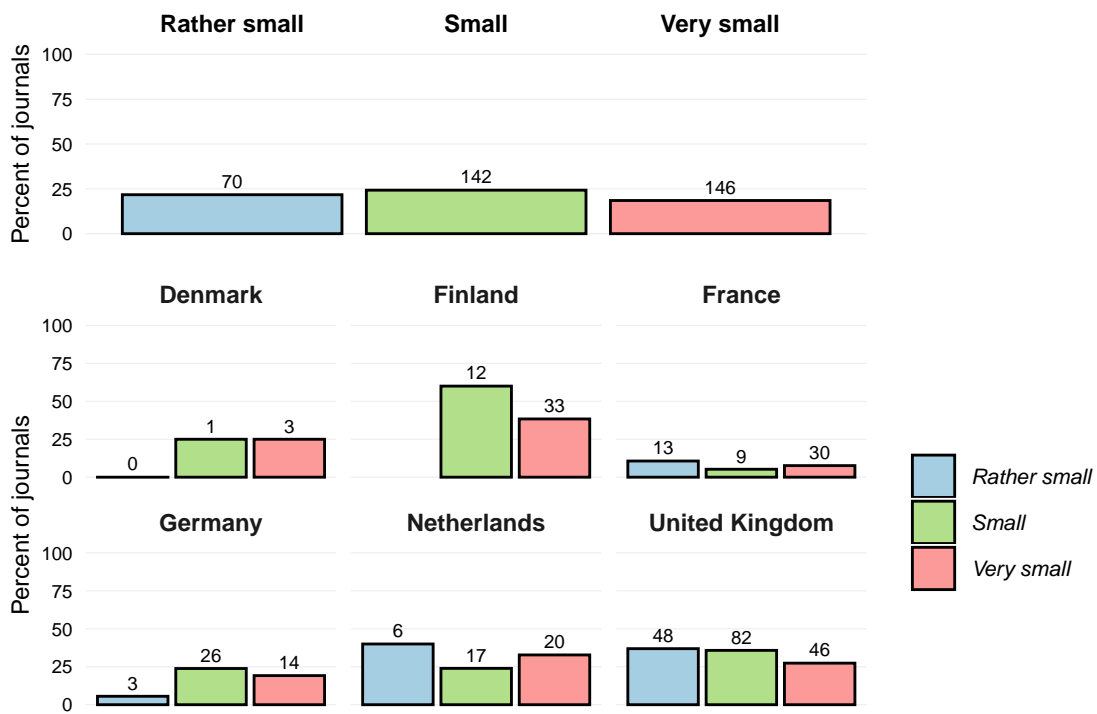


Figure 6: Number and percentage of journals in Sherpa Romeo by publisher category, and country.

In Figure 7 we show the number and percentage of publishers by OA status and category and country based on the status in Unpaywall. Publishers were assigned to each OA status on the basis of the status of their journals derived from the status of the individual documents in each journal. As such, if all documents were classified as open or closed in all of the publishers' journals, then the publisher was assigned a status of open or closed respectively. Consequently, the majority of publishers were mixed, as their documents and subsequently journals were a mixture of open, closed, or OA variations (green, bronze, hybrid).

Very small publishers were more likely to be classified as open or closed – rather than mixed – as their status was based upon only one journal in comparison to the 2-10 journals produced by small and rather small publishers. Publishers with the status “Unknown” are those that had no DOIs to be searched in Unpaywall. The majority of each country's publishers were classified as mixed, except Finland, for which nearly 60% of publishers published OA. Smaller percentages of 25-33% of publishers in Denmark and the Netherlands, and 14-18% in France, Germany, and the United Kingdom also published OA. Germany had the highest percentage of closed publishers at 12%.

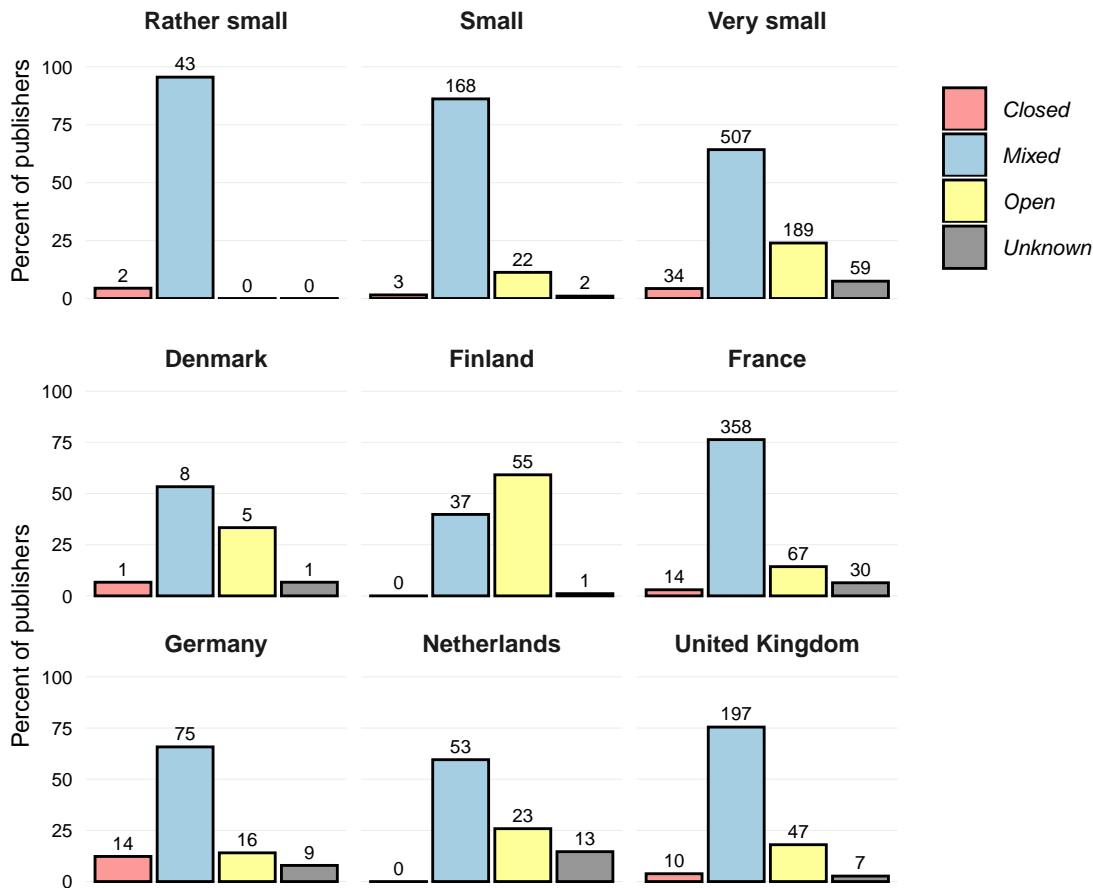


Figure 7: Number and percentage of publishers by OA status derived from Unpaywall, publisher category, and country.

The complementary perspective of the number and percentage of journals classified by OA status by publisher category and country is shown in Figure 8. This perspective highlights that, for instance, rather small publishers do use open publishing pathways, although for only a quarter of their journals.

Further, the distribution of journals across the OA categories was quite similar between publisher categories, with approximately 60% of journals using mixed models, around 25% using open models, and the minority of journals published as closed access. However, the use of closed publishing declined by publisher size, with 10%, 7%, and 4% of journals or rather small, small, and very small publishers using closed models respectively.

Further, like Finland, a slight majority of journals were published OA by Danish publishers. However, they are combined with mixed models such that the majority of Danish publishers are considered mixed. Considering the higher percentages of journals published OA in comparison to the percentage of publishers classified as OA, this appears to be a common trend among publishers in all six countries to publish a portion of their journals OA, alongside journals using mixed or closed pathways. Notably, French and German publishers published their journals OA less often than the other countries, and German publishers were more likely to use closed models (16% versus 1-7%). The higher rate of journals of unknown status from French publishers reflects the introduction via the supplementary lists of journals without DOIs.

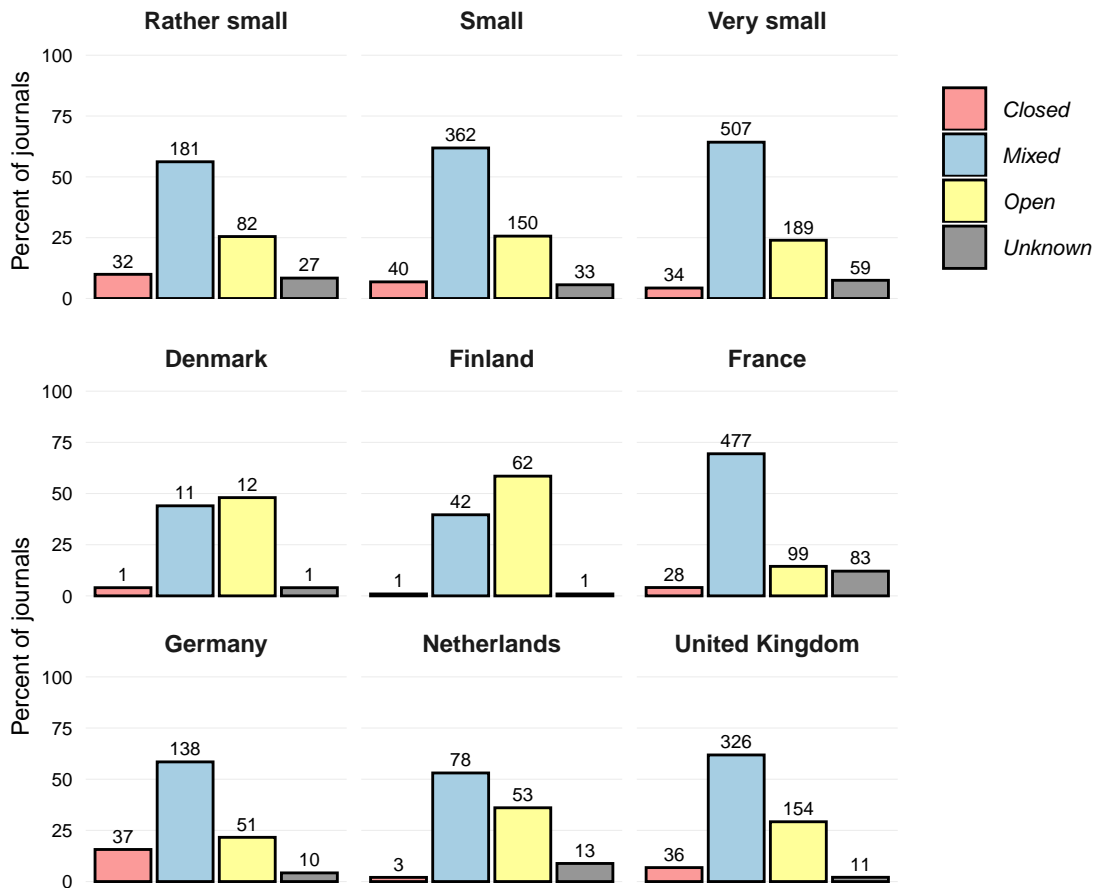


Figure 8: Number and percentage of journals by OA status derived from Unpaywall, publisher category, and country.

We show in Figure 9 an aggregated view of the OA status of all small publishers in the KE countries based on data from Unpaywall and the DOAJ. We exclude here data from Sherpa Romeo as coverage in this resource is not indicative of an OA status and the Crossref data were too incomplete to facilitate

a meaningful analysis. As such, we show here the number and percentage of all small publishers – that is, publishers in all three groups and from all six countries – and their journals that were of closed, open, mixed, or unknown status in Unpaywall, or that were indexed in the DOAJ and are thus open. The classification of open or closed was based on all documents/journals being classified as open or closed in Unpaywall, otherwise the journal or publisher was classified as mixed.

Overall, we see here that, using this classification, the majority of both journals (61.9%) and publishers (69.8%) were mixed, and only a small percentage of both journals (6.3%) and publishers (3.8%) used a closed publishing model. The percentage of journals and publishers classified as open were similar between Unpaywall and DOAJ: 24.8% and 23.8% of journals and 20.5% and 31.3% of publishers, respectively. The close alignment between the DOAJ and Unpaywall at the journal level suggests that the classification of journals as open via Unpaywall is likely valid. The slightly higher number of journals in Unpaywall may stem from, for instance, open journals that have not applied for indexation in the DOAJ. The differences in the number and percentage of open publishers between the DOAJ and Unpaywall, however, reflect again this tendency for publishers to publish a proportion but not all of their journals as OA. Under the classification used here, such publishers are considered mixed based on Unpaywall data, but open for DOAJ.

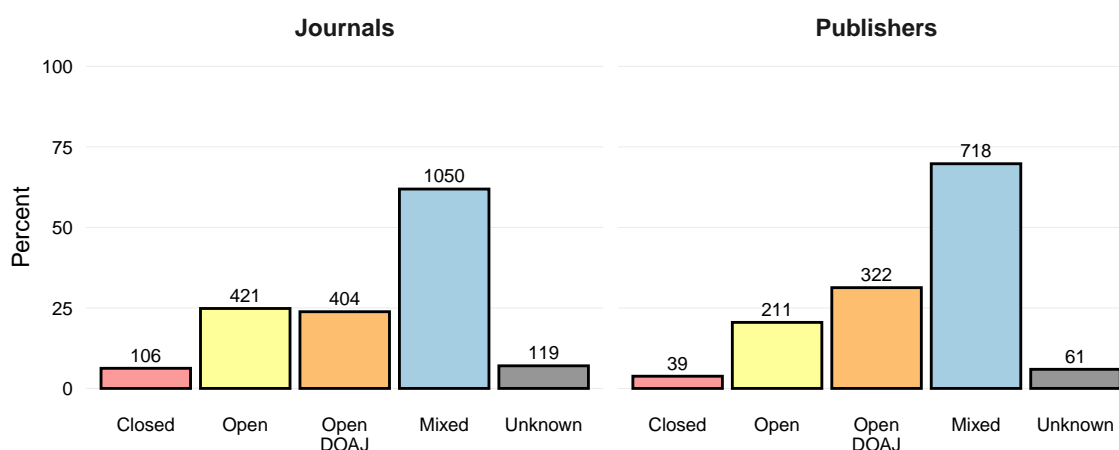


Figure 9: The overall number and percentage of small publishers and their journals by OA status in DOAJ and Unpaywall.

To provide a sense of the composition of mixed journals, we show in Figure 10 the distribution of the percentage of documents classified in mixed journals as gold open, closed, or an OA variation other than gold, such as green or bronze. For example, a mixed journal may be comprised of 20% documents classified as open, 30% classified as closed, and the remaining 50% as OA variations. In Figure 10, a value is provided for each OA type in each band of percentages. That is, closed documents comprised 0-10% of just under 600 mixed journals, OA variation was 0-10% of 400 journals, and OA documents comprised less than 10% of 800 mixed journals. As journals with 100% of closed or open documents were classified accordingly, journals in the 90-100% bin are those with up to 99% open or closed documents.

We see by the two peaks in the 0-10% and 90-100% bins that the majority of mixed journals were skewed toward one type of OA status – most often OA variations or closed – with a much smaller percentage of documents classified as other types. Only a minority of mixed journals had similar proportions of documents of different OA types. Open access was particularly polarised, with OA

documents making up typically less than 10%, or in fewer cases, more than 90% of journals' content. Overall then, mixed journals tend to be a combination of documents published as closed and OA variations, but with a skew toward one type of OA.

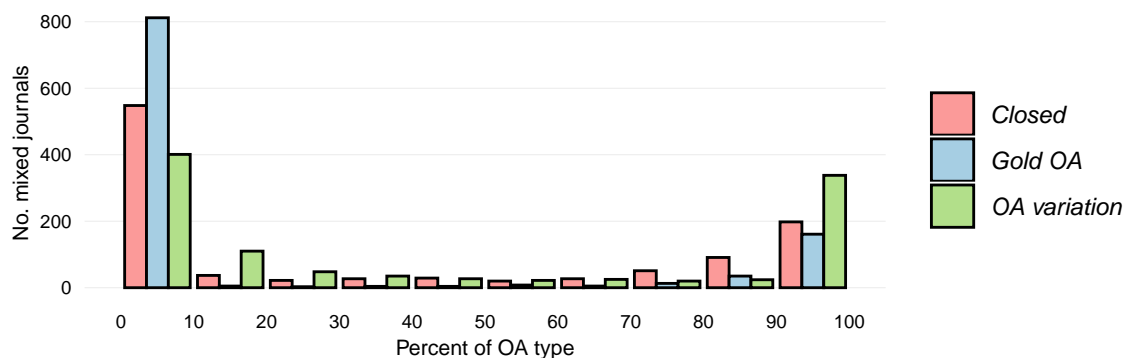


Figure 10: Distribution of the percentage of open, closed, and OA variation documents in mixed journals based on OA information in Unpaywall.

We can then delve further into the publishing conditions of journals based on the information available in Sherpa Romeo, including who retains copyright ownership, publishing conditions, and whether additional fees are required to publish OA. However, generalisations of these findings to broader publisher categories or countries should be carefully considered as the journals present in Sherpa Romeo represented less than a quarter of small publishers' journals.

Some percentages do not sum to 100% within publisher groups because there were occasionally multiple pathways to publishing with different conditions. Table S4 in the Supplementary Material provides the following information for the submitted, accepted, and published versions of articles in all journals: whether OA is prohibited, the licenses used, copyright ownership, the requirement for additional OA fees, the publishing conditions, embargo periods, and the publishing locations permitted (e.g. journal, institutional repository).

Figure 11 shows the percentage of journals by publisher category and country that award copyright ownership to the authors, publishers, and for which it was not provided. For the majority of journals in each publisher category and country, the author retains copyright ownership for published articles. However, copyright information was not known for between 25% and 50% of journals in all groups and countries. Hence these data did not strongly differentiate between groups.

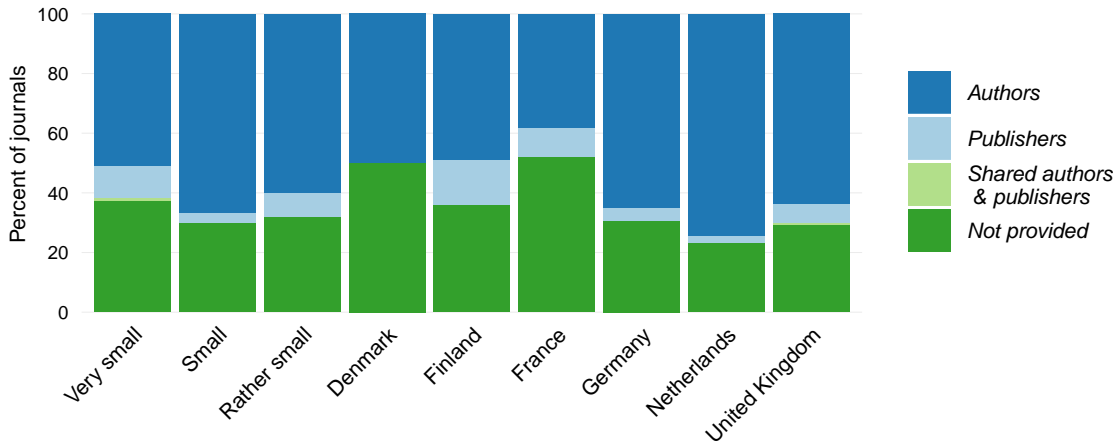


Figure 11: The percentage of journals by copyright owner, publisher category, and country, based on Sherpa Romeo data.

The requirement to pay additional fees for publishing OA by publisher category and country are shown in Figure 12. In interpreting this information, it is important to note that OA fees are defined by Sherpa Romeo as “a fee that is paid *in addition to other publication fees*, in order to make an article open access immediately.”³⁰ The “additional OA fee” indicator will typically be “no” for OA journals, as they do not require additional fees for publishing OA. However, the additional OA fee indicator is often left blank for OA journals as any additional fees cannot be differentiated from the Article Processing Charges (APC) used by some OA journals (personal communication, 29 June 2022). Consequently, only a “yes/no” response definitively identifies whether additional charges for publishing OA apply. Across all categories and countries, the majority of journals did not require an additional OA fee. Journals from small publishers and German and British publishers imposed additional OA fees slightly more often than the other publisher categories and countries. However, the level of missing data and similarities between groups meant these data were not particularly helpful in distinguishing between publishers.

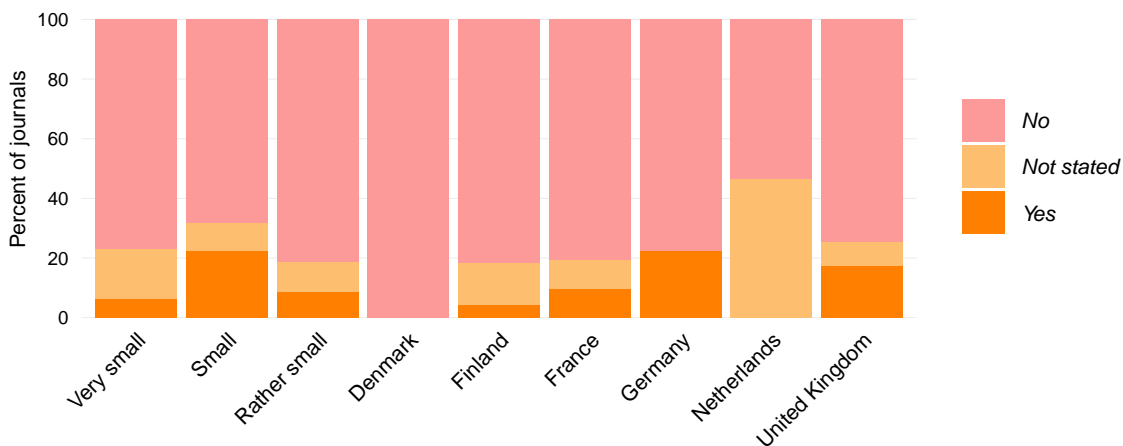


Figure 12: The percentage of journals by additional OA fees status, publisher category, and country, based on Sherpa Romeo data.

³⁰<https://v2.sherpa.ac.uk/romeo/about.html>

Table 4: Number and percent of journals by publishing condition and publisher category, based on Sherpa Romeo data.

Conditions	Category	No. journals	Percent
Must link to published article with DOI/set statement	Rather small	5	7.1
Must link to published article with DOI/set statement	Small	45	31.7
Must link to published article with DOI/set statement	Very small	42	28.8
No conditions noted	Rather small	14	20.0
No conditions noted	Small	46	32.4
No conditions noted	Very small	65	44.5
Published source/copyright must be acknowledged	Rather small	69	98.6
Published source/copyright must be acknowledged	Small	90	63.4
Published source/copyright must be acknowledged	Very small	74	50.7

The number and percentage of journals by publisher category stipulating each publishing condition for published article versions are shown in Table 4. Publishing conditions were particularly uniform with either no conditions noted, or two other conditions required: the published source or copyright owner must be acknowledged with a citation and/or set statement, or that the article must be linked to with a DOI or set statement. The business process was more often not completely defined in detail by very small publishers (44% of journals had no conditions) than small or rather small publishers (32% and 20%, respectively). Nearly all rather small publishers required the published source or copyright owner was acknowledged, while more small and very small publishers required a link to the published article via DOI or set statement. A small number of journals had other conditions, and these can be seen in Table S4 of the Supplementary Material.

Conclusions and considerations

It appears that Crossref provides relatively limited information regarding the OA status of publishers. The provision of a URL to license conditions is potentially useful information. However, the high level of missingness of these data and the requirement that the conditions must be examined to translate them into an OA status reduces the applicability of these data.

Coverage of the publishers and journals in the DOAJ and Sherpa Romeo was not high, at around 22% of journals and publishers in Sherpa Romeo and 24% of journals and 31% of publishers in the DOAJ. This level of coverage in the DOAJ is in itself informative, as the similar estimates of OA journals between the DOAJ and the Unpaywall-derived classification suggests that the DOAJ captured the majority of OA journals within the sample. From the DOAJ data, we observed that rather small publishers were slightly better represented than small or very small publishers. However, this pattern was reversed when considering the journal-oriented perspective, which aligns with the tendency observed via the Unpaywall data for publishers to publish some but not all of their journals as OA. Coverage in DOAJ was similar between publisher countries.

In contrast, there are theoretically no limitations on inclusion in Sherpa Romeo and so its modest coverage of the small publishers is less informative. Sherpa Romeo is a promising source of additional information about the conditions and associated details of publishers' policies. However, the incompleteness of these data sometimes limited its utility in the current study. Some variation in coverage was observed, with very small, French, and German publishers less well represented than other publisher categories and countries.

As expected due to the interdependencies between Crossref and Unpaywall, coverage in Unpaywall of the small publishers was essentially complete. The OA status classification of each document was also complete, making this source invaluable to the OA status classification of journals and publishers. Based on these data, the majority of publishers in each category and country – except Finland – were classified as mixed publishers. Publishers in Finland published predominantly OA. Approximately a quarter of very small publishers, and publishers from Denmark and the Netherlands were also OA, while very few publishers were entirely closed.

Notably, the Unpaywall-based OA status classification used in the current study is sensitive to the thresholds used. Journals were classified as mixed when they published less than 100% of documents as open or closed. This threshold was selected on the basis that it is expected that OA journals publish all documents OA and thus any percentage less than 100 would indicate mixed publishing practises. However, less stringent thresholds could also be used to define the status of journals and publishers. For instance, in the comparison of publishers of different OA types in Section 5 of this report, we used a 90% threshold for classification based on the distribution of documents shown in Figure 10. The inclusion in Table S2 of the number and percentage of documents per journal of each status facilitates additional studies using differing thresholds, if desired.

Unfortunately, the helpful extension of the coverage of French and Dutch journals by the supplementary lists meant that DOIs were not available for these journals and they had to be excluded from the analysis in Unpaywall. While beyond the scope of the current studies, future studies could supplement the dataset with information from additional DOI repositories to facilitate the inclusion of such publishers.

4. Characteristics of small publishers

The aim of this Section of the study was to explore the feasibility of determining other relevant characteristics of small publishers via the ISSN Portal. The ISSN Portal was the most appropriate data source for assessing these characteristics as – being an integral data source in defining the sample – it had complete coverage of the sample and also held data for the characteristics of interest. These characteristics include the language(s) accepted by the journals, their disciplines, changes over time in the publisher of the journal, the lifespans of publishers, the frequency with which journals are published, and any discontinued or newly created journals within the 3-year period of analysis. Such information can help to establish a broader picture of the landscape of small publishers and a profile of the characteristics of publishers in each category and country.

However, the characteristics to be examined necessitated commercial access to the “Complete ISSN data”, as described by the ISSN Portal³¹. Consequently, these data are not publically available and cannot be shared at the journal-level as provided in the Supplementary Material for the previous data sources. As such, we provide here several presentations of the data and further disaggregations in Tables S5-S11 of the Supplementary Material. However, these data should be interpreted carefully due to the often small sample sizes involved. The method used to retrieve these data from the ISSN Portal are described in the Methods section of Section 2.

³¹<https://portal.issn.org/faq2>

Results

Publishing languages

The number and percentage of journals that accepted manuscripts in selected languages are shown in Figure 13, disaggregated by publisher category, and country. Supplementary Table S5 shows the number of journals published in these languages for each country's publishers by category. Journals have been counted once toward each language it accepts. As such, the counts may exceed those presented in Table 2 and percentages may sum to more than 100%. *Multiple* includes journals that publish in multiple languages, but for which the specific languages were not stated.

	Very small	Small	Rather small	Denmark	Finland	France	Germany	Netherlands	UK
Danish	3, 0.4%		4, 1.2%	7, 28%					
Dutch	19, 2.4%	8, 1.4%						27, 18.4%	
English	370, 46.9%	347, 58.6%	161, 50%	17, 68%	54, 50.9%	153, 22.3%	95, 40.3%	108, 73.5%	477, 90.5%
Finnish	59, 7.5%	7, 1.2%			66, 62.3%		1, 0.4%		
Swedish	26, 3.3%	1, 0.2%		1, 4%	26, 24.5%		1, 0.4%		
French	356, 45.1%	145, 24.5%	98, 30.4%		2, 1.9%	575, 83.7%	12, 5.1%	4, 2.7%	10, 1.9%
German	38, 4.8%	40, 6.8%	34, 10.6%		3, 2.8%	11, 1.6%	92, 39%	3, 2%	6, 1.1%
Multiple	129, 16.3%	65, 11%	46, 14.3%	1, 4%	13, 12.3%	150, 21.8%	58, 24.6%	9, 6.1%	13, 2.5%
Other	63, 8%	48, 8.1%	41, 12.7%	1, 4%	2, 1.9%	88, 12.8%	22, 9.3%	11, 7.5%	30, 5.7%
Unspecified		20, 3.4%	21, 6.5%			5, 0.7%	6, 2.5%		30, 5.7%

Figure 13: Number and percentage of journals in selected languages by publisher category, and country, based on ISSN Portal data.

Approximately half of publishers from each category published their journals in English, making it the most prevalent publishing language. French was also popular for all publishers, but particularly very small publishers, who published 45% of their journals in French. These trends are likely influenced by the relatively large proportions of British and French publishers in the sample, the majority of whom published in English (90.5%) and French (83.7%) respectively, in addition to English being the lingua franca of international discourse in the global academic system.

Similar to French publishers, whose dominant publishing language was the local one, Finnish publishers most often published their journals in Finnish (62.3%), with smaller percentages in English

(50.9%) and Swedish (24.5%). Conversely, English was the predominant language used by journals from Danish and Dutch publishers, with 20-30% published in the local language. German publishers were evenly split between German and English. In comparison, publications in languages such as Chinese, French, German, Spanish and Italian have been found to each comprise less than 6% of content in Scopus and less than 2% of content in WoS [17]. This suggests that small publishers are much more likely than larger publishers to publish journals in local languages, either solely or in addition to in English.

Publishers' choices of the languages accepted for their journals may reflect the size of the population of speakers, the journal's subject matter, and the intended audience of the journal. For instance, nationally-oriented or arts and humanities journals may be better suited to being published in local languages than journals addressing globally-relevant research topics in the physical and medical sciences.

Subjects

The number and percentage of journals attributed to each subject, disaggregated by publisher category, and country, is shown in Figure 14. These data are drawn from the Universal Decimal Classification system (UDC)³² used in the ISSN Portal. This classification consists of nine broad categories. However, its utility for observing trends in academic journal classification is limited by the distribution of related topics across several categories and inclusion of diverse topics in one category. As such, we have mapped the UDC to the OECD's Fields of Science classification³³, which uses six broad, distinct categories suitable for the analysis of journal classifications.

The Finnish National Library, which supplies information about Finnish journals to the ISSN Centre, kindly provided the mappings between the UDC codes provided in the ISSN Portal and the UDC categories (personal communication, 22 July 2022). The mapping between the UDC and the OECD classification is available in Supplementary Table S6. Journals could be attributed to more than one subject and as such, the counts of journals may exceed those presented in Table 2 and percentages may sum to more than 100%. The allocation of journals to subjects by both publisher category and country is shown in Supplementary Table S7.

The largest share of most publishers' journals were in the humanities and social sciences, accounting for between 26% and 48% of journals. Finnish, German, and French publishers were particularly focused on the humanities and social sciences, with nearly half of their journals pertaining to these topics. Dutch publishers produced a diverse array of journals as, while the majority were in the social sciences and humanities, a further 20% each were classified to the natural sciences and medical sciences. Notably, a quarter of British journals were unclassified. The overall patterns observed here deviate from the subject-based composition of WoS and Scopus, where natural science and medical science publications comprise the largest portions of these databases[17]. Conversely, there is a much stronger orientation toward the humanities and social sciences amongst these small publishers.

³²<https://finto.fi/udcs/en/page/078887>

³³<https://www.oecd.org/science/inno/38235147.pdf>

	Very small	Small	Rather small	Denmark	Finland	France	Germany	Netherlands	UK
Natural sciences	87, 11%	94, 16.1%	27, 8.4%	7, 28%	21, 19.8%	64, 9.3%	35, 14.8%	30, 20.4%	57, 10.8%
Engineering & tech.	64, 8.1%	58, 9.9%	16, 5%	5, 20%	10, 9.4%	49, 7.1%	22, 9.3%	8, 5.4%	48, 9.1%
Medical sciences	70, 8.9%	56, 9.6%	43, 13.4%	3, 12%	6, 5.7%	37, 5.4%	18, 7.6%	27, 18.4%	81, 15.4%
Agricultural sciences	12, 1.5%	14, 2.4%	3, 0.9%		5, 4.7%	9, 1.3%	4, 1.7%	2, 1.4%	9, 1.7%
Social sciences	311, 39.4%	179, 30.6%	102, 31.7%	4, 16%	37, 34.9%	316, 46%	61, 25.8%	49, 33.3%	135, 25.6%
Humanities	302, 38.3%	165, 28.2%	106, 32.9%	7, 28%	51, 48.1%	309, 45%	97, 41.1%	38, 25.9%	74, 14%
Not available	59, 7.5%	76, 13%	49, 15.2%		5, 4.7%	20, 2.9%	14, 5.9%	14, 9.5%	133, 25.2%

Figure 14: Number and percentage of journals published by subject and publisher category, and country, based on ISSN Portal data.

Lifespan of publishers

The distribution of the ages of publishers by category and country are shown in Figure 15, and by both country and category in Supplementary Table S8. A publisher's age was calculated by deducting the earliest starting year of any of the publishers' journals from the year 2022. The continuous horizontal line represents the full distribution of the publishers' ages, while the three vertical lines constituting the box indicate the lower quartile (25%), median, and upper quartiles (75%), respectively. The small grey points show all the individual publishers' ages, and the larger black points are considered outliers on the basis that the age exceeds the upper quartile by more than 1.5 times the inter-quartile range.

The median ages of publishers in all groups fell between 11 and 25 years. However, the oldest publishers began publishing more than 150 years ago. Manual checks were conducted on a sample of the oldest publishers to confirm the starting years were accurate. Very small publishers tended to be slightly younger (median = 18 years) than small and rather small publishers (23 and 25 years). Danish and British publishers were also typically younger than other countries (median = 11 and 12 years), while French publishers were older with a median of 23 years. Nearly all countries had at least one publisher established more than 130 years ago, while Denmark's oldest small publisher began publishing 54 years ago.

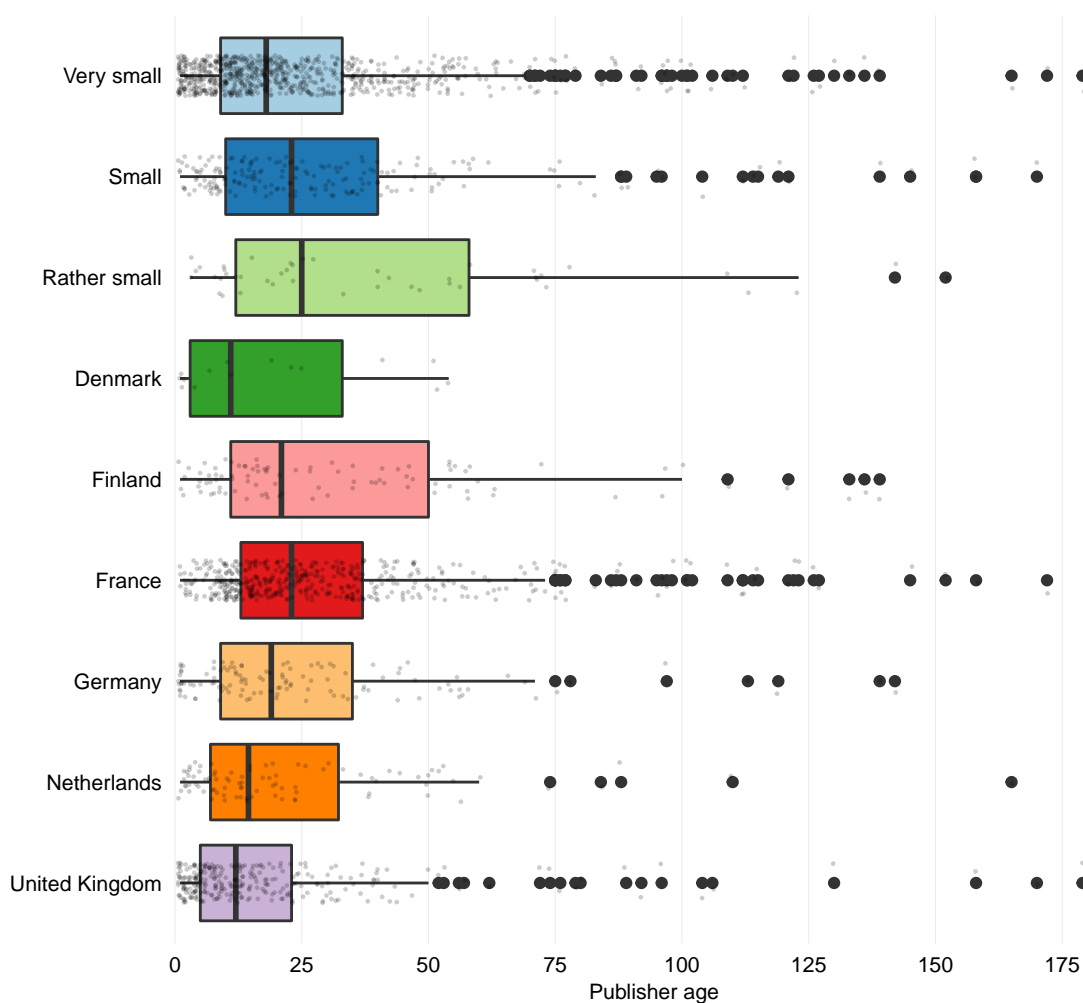


Figure 15: Age of publisher in 2022 based on earliest journal by publisher category, and country, based on ISSN Portal data.

Changes in publisher

The number and percentage of journals for which the publisher changed are shown in Figure 16 by the category and country of the current publisher, and by both variables in Supplementary Table S9. Changes in a journal’s publisher were determined by identifying differences in the names of the publishers between the earliest and latest publication events. However, such changes could not be determined for a number of journals as the publisher at the earliest publication event was not recorded. This particularly affected journals from Danish and German publishers. The large majority of journals retained the same publisher over time; across groups, it was typical for 8-10% of journals to change publishers. The exception was Finland, where only 3.8% of journals had a different publisher now than when the journal was first published. A lower than average percentage of German journals also changed publishers. However, these data may be unreliable as 40% of data pertaining to the initial publisher were missing.

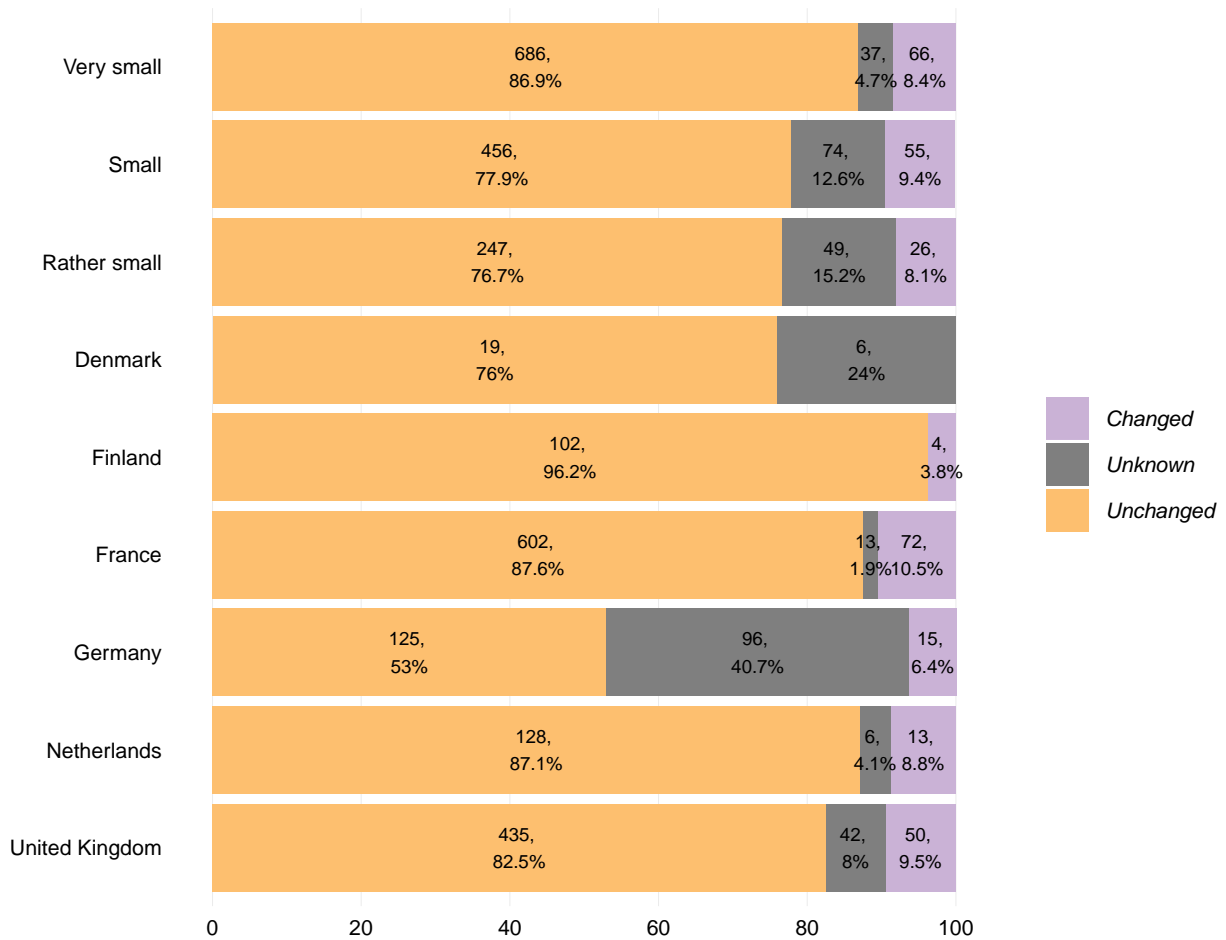


Figure 16: The number and percentage of journals and whether they changed publishers over time, by publisher category, and country, based on ISSN Portal data.

Status of journals

The number and percentage of journals that were continued or the publication of which was started or ended during the 2019-2021 period by publisher category and country are shown in Figure 17 and Supplementary Table S10. Overall, small publishers appeared to be quite well established, with the majority of journals continued throughout this period, a notable percentage added, and only a minority discontinued. French publishers were particularly stable; the large majority of journals were continued and only 5% of journals were introduced and 3% ended during this period. In contrast, Danish and Dutch publishers appeared to be experiencing a period of growth, with over 20% of their journals introduced during these three years and 4-5% discontinued.

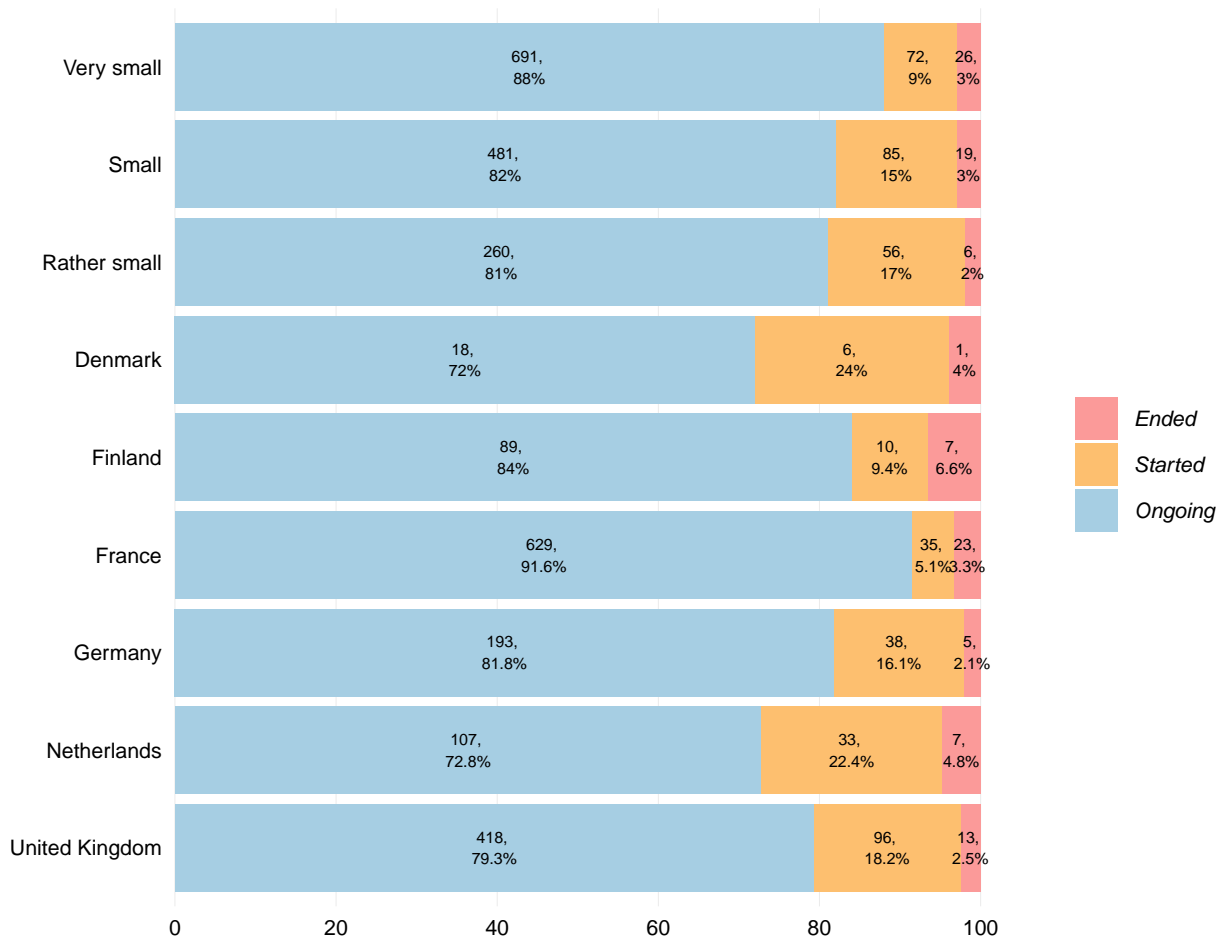


Figure 17: The number and percentage of journals that were continued, started, or ended during 2019-2021 by publisher category, and country, based on ISSN Portal data.

Accrual periodicity

In Figure 18 and Table S11, the number and percentage of journals published by particular accrual periodicities are shown, disaggregated by publisher category and country. Broadly, small publishers of all groups tended to opt for two publishing schedules, quarterly and semi-annually, with a somewhat smaller percentage publishing annually. As might be expected with small publishers due to costs and collecting content, few journals were published with any frequency shorter than bimonthly. Seven Danish publishers (28%) also published on a completely irregular schedule, which was much higher than other groups. However, this perhaps extends from the small sample of Danish publishers. Also, nearly a fifth of German publishers published on schedule other than those described. However, the details of these schedules were not stated.

	Very small	Small	Rather small	Denmark	Finland	France	Germany	Netherlands	UK
Continuous	6, 0.8%	5, 0.9%		1, 4%	1, 0.9%	2, 0.3%		5, 3.4%	2, 0.4%
Biweekly	2, 0.3%					1, 0.1%		1, 0.7%	
Weekly	1, 0.1%	1, 0.2%					1, 0.4%	1, 0.7%	
Semi-monthly	2, 0.3%					1, 0.1%			1, 0.2%
Monthly	25, 3.2%	21, 3.6%	6, 1.9%		1, 0.9%	21, 3.1%	6, 2.5%	10, 6.8%	15, 2.8%
Bimonthly	38, 4.8%	40, 6.8%	16, 5%	1, 4%	3, 2.8%	28, 4.1%	19, 8.1%	5, 3.4%	40, 7.6%
Quarterly	186, 23.6%	132, 22.6%	54, 16.8%	2, 8%	35, 33%	145, 21.1%	37, 15.7%	44, 29.9%	120, 22.8%
Three times a year	43, 5.4%	49, 8.4%	10, 3.1%	1, 4%	1, 0.9%	36, 5.2%	10, 4.2%	10, 6.8%	47, 8.9%
Semiannual	221, 28%	151, 25.8%	99, 30.7%	10, 40%	29, 27.4%	205, 29.8%	58, 24.6%	48, 32.7%	130, 24.7%
Annual	115, 14.6%	81, 13.8%	62, 19.3%	2, 8%	24, 22.6%	118, 17.2%	34, 14.4%	9, 6.1%	75, 14.2%
Biennial	4, 0.5%	2, 0.3%	3, 0.9%			5, 0.7%		1, 0.7%	3, 0.6%
Other	38, 4.8%	22, 3.8%	15, 4.7%	1, 4%	1, 0.9%	26, 3.8%	44, 18.6%	2, 1.4%	2, 0.4%
Completely irregular	54, 6.8%	35, 6%	21, 6.5%	7, 28%	11, 10.4%	20, 2.9%	11, 4.7%	11, 7.5%	50, 9.5%
Not specified	54, 6.8%	46, 7.9%	36, 11.2%			79, 11.5%	16, 6.8%		42, 8%

Figure 18: The number and percentage of journals published by accrual periodicity type, publisher category, and country, based on ISSN Portal data.

Conclusions and considerations

This Section of the study examined several characteristics of small publishers to aid in developing a profile of the publishers of each category and country. For instance, it emerged that, as the lingua franca of academia, English was a prominent publishing language for all publisher categories and in all countries. Typically approximately half of journals were published in English. However, this was understandably higher for British publishers at 90% of their journals and Dutch and Danish publishers also produced more journals in English than their local languages. Conversely, French and Finnish publishers were more likely to publish in local languages than English. The use of languages other than English by small publishers far exceeded the levels observed for publishers in Scopus and WoS, highlighting the particular discourse upheld by these small publishers.

The small publishers also tended to be more oriented toward the humanities and social sciences than the larger publishers in commercial databases. In particular, nearly half of Finnish, French, and German journals pertained to the social science and humanities. In terms of publishers' ages, Danish, British, and very small publishers tended to be younger than other publishers, while French publishers were older. Journals were also relatively unlikely to change ownership over time, with around 8-10% of journals being published by a different publisher now than when they were initially introduced. Finnish journals, in particular, were least likely to change publishers. Small publishers appear to be experiencing growth; all categories and countries introduced more than 5% – and typically closer to 15% – of their journals in the last three years, while discontinuing only around 3% over the same period. The vast majority of journals were continued throughout the period. All groups most often published their journals on a quarterly or semi-annual basis, with a smaller percentage publishing annually. From these characteristics we can observe the distinctions and similarities between the profiles of small publishers of varying sizes and between countries.

However, it should be considered that the ISSN Portal is provided as an information tool and not as a database for journal or publisher analytics. As such, there is occasionally a level of missingness in the data that must be considered, such as in the case of subject classifications for British journals, and original publisher information for especially German journals. Further, we reiterate again that we have applied here the lens of Crossref and supplemented this with data from the ISSN Portal. Consequently, there may be a level of variability introduced into the results observed here due to differences between the two sources.

5. Contrasting small publishers that have and have not transitioned to OA

This last Section of the study addresses the third research question regarding what differences can be observed between small publishers that have and have not begun to transition to OA. This analysis identifies specific characteristics of publishers of different OA types that may indicate aspects of the transition that require particular support. This information could be useful to the KE in developing recommendations for the support and continuation of small publishers during the transition to OA.

Methods

As previously described, we allocated journals and publishers a status of either open, closed, or mixed on the basis of the classification of the journals' individual documents in Unpaywall. However, the

distribution of documents within mixed journals – as shown in Figure 10 – indicated that journals tended to be comprised of documents of primarily one type. Consequently, we decreased the threshold for a journal to be considered open or closed to 90% for this analysis. As such, a journal was considered open or closed when more than 90% of its documents were classified as such, otherwise it was considered mixed. As before, publishers were classified on the basis of their journals, with publishers being categorised as open or closed when all of their journals were classified as such. Publishers that published only mixed journals or a combination of OA types were classified as mixed. This sought to better reflect the primary classification of open and closed publishers' journals to improve the reliability and validity of results obtained via quantitative comparison of the publishers in each group. We then compared the publisher groups on their characteristics as determined in the previous components of the study. Please note, OA data for 66 journals from 61 publishers (2 small, 61 very small) could not be obtained and these journals/publishers are excluded from this analysis. As such, the counts of journals and publishers differ between this and previous Sections.

Results

We show in Figure 19 the number and percentage of publishers classified as open, closed, mixed, or unknown status based on a 90% document-level threshold. The effect of the 90% threshold in reassigning publishers can be observed by comparing Figures 7 and 19. As with using a 100% threshold classification, the majority of publishers in all categories and countries were mixed. The exception was Finland's publishers, over 60% of which publish their journals OA. Laakso and Multas [9] similarly observed this tendency for Finland to have a higher percentage of OA small-to-mid-sized publishers (41%) than other KE countries (24%-28%). The Netherlands also had nearly equal proportions of open and mixed publishers. In contrast, there was a higher percentage of closed publishers in Germany and France (18-23%) than the other countries (1-12%). The OA status of very small publishers was quite different to small and rather small publishers, with a lower proportion of mixed publishers and greater proportions of OA and closed publishers. However, very small publishers' OA status was defined on the basis of their one journal, compared to multiple journals for the other publishers.

Publishers of unknown status have been removed from subsequent analyses as their uncertain status means their relation to open, closed and mixed publishers cannot reliably be determined, and their unknown status may be the only characteristic connecting the journals, making a collective and contrasting assessment unreliable. Further, due to the small sample sizes and further disaggregations introduced by additional characteristics, we refrain from contrasting publisher groups at the category and country levels here and present data in the remaining analyses by publisher OA status. Tables S12 to S24 in the Supplementary Material present data disaggregated by both country and publisher category. However, care should be taken when interpreting these data due to the very small sample sizes frequently involved.

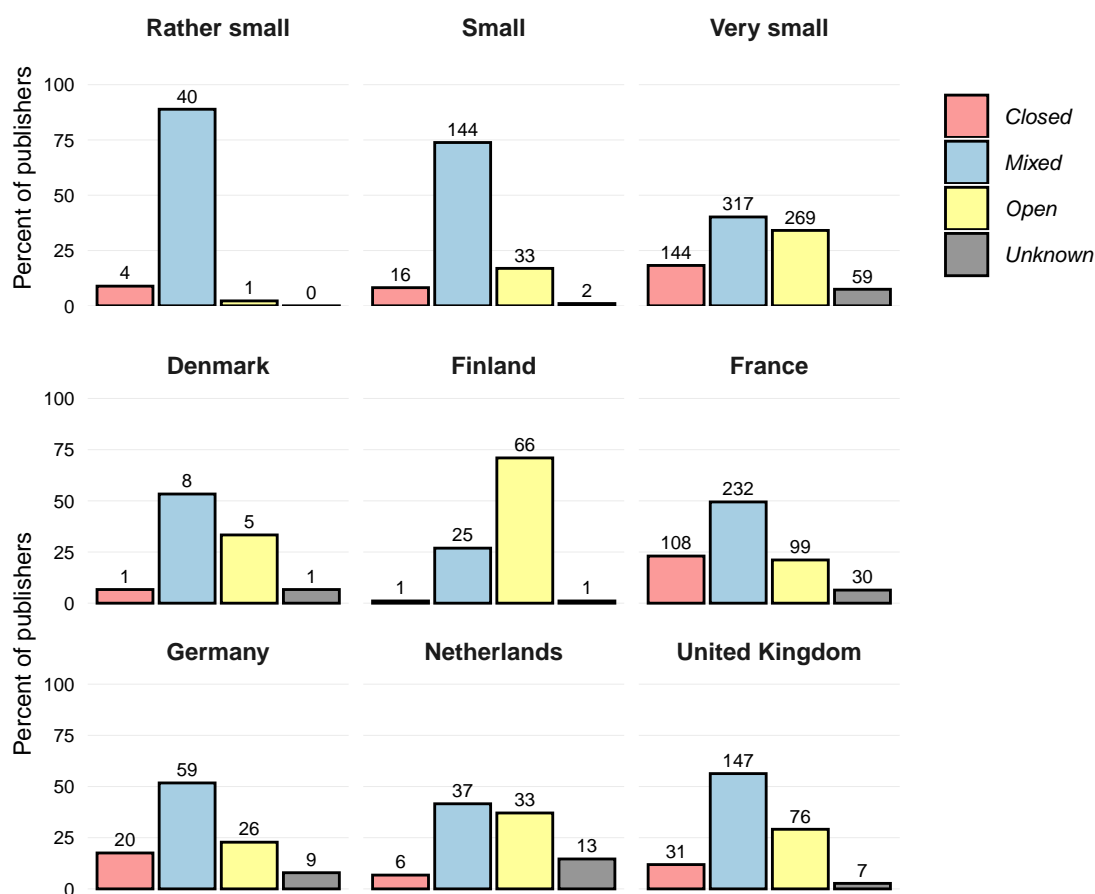


Figure 19: Number and percentage of publishers by OA status derived from Unpaywall using 90% document threshold by publisher category, and country.

The number and percentage of journals accepting publications in selected languages is shown in Figure 20 by publishers' OA status. Here, "local languages" refers to the non-English languages local to the KE countries; Danish, Dutch, Finnish, Swedish, French, and German. Multiple languages refers to cases in which the journal accepts several languages but the specific languages accepted were not recorded. Journals were counted toward each language they accept and as such, the percentages may not sum to 100% within each OA status group.

It becomes evident that open publishers are more likely to produce their journals in English (66.9%), than are mixed (52.4%) or closed publishers (28.6%). This is likely influenced by the large number of very small French publishers included in the closed sample, which produce their journals primarily in French (Figure 13). The emphasis on English by open publishers goes beyond the influence of British publishers – which comprise another large portion of all publishers but only approximately a third of open publishers – and likely reflects an intention by these publishers to appeal to an international audience. Notably, however, accepting a language does not necessarily reflect the extent of its use within the journal. Further studies may wish to examine the composition of journals by their publishing language using, for instance, the document-level language data in Crossref for an additional perspective of language use by small publishers.

	English	Local languages	Multiple languages	Not available	Other language
Open publishers	243, 66.9%	150, 41.3%	69, 19%		25, 6.9%
Closed publishers	61, 28.6%	139, 65.3%	27, 12.7%	12, 5.6%	13, 6.1%
Mixed publishers	552, 52.4%	456, 43.3%	134, 12.7%	28, 2.7%	83, 7.9%

Figure 20: The number and percentage of journals published in selected languages by publishers’ OA status.

In Figure 21 the number and percentage of journals in each subject area of the OECD’s Fields of Science classification are shown by publishers’ OA status. Journals were counted once toward every subject area they were classified to and so percentages may exceed 100% when summed per publisher group. As in Figure 14, small publishers had a strong focus on social science and humanities topics. Approximately a third each of open and mixed publishers’ journals were in these subject areas. There was less emphasis on the humanities from closed publishers, although 27.7% of their journals still addressed humanities topics. Further, open publishers had a slightly higher percentage of journals in the natural sciences than the other publishers, while closed publishers’ content was more often in engineering and technology, which might differ from the internationally-focused natural sciences to a stronger focus on local applications.

	Open publishers	Closed publishers	Mixed publishers
Natural sciences	54, 14.9%	24, 11.3%	123, 11.7%
Engineering & tech.	33, 9.1%	26, 12.2%	75, 7.1%
Medical sciences	38, 10.5%	19, 8.9%	109, 10.3%
Agricultural sciences	6, 1.7%	7, 3.3%	16, 1.5%
Social sciences	129, 35.5%	77, 36.2%	361, 34.3%
Humanities	139, 38.3%	59, 27.7%	348, 33%
Not available	27, 7.4%	26, 12.2%	127, 12%

Figure 21: The number and percentage of journals by subject area and publishers’ OA status

The frequency with which publishers of different OA statuses published their journals is shown in Figure 22. Somewhat different publishing patterns emerged between publishers, with OA and mixed publishers releasing two-thirds of their journals on a semiannual, quarterly or annual publishing schedule. Conversely, closed publishers tended to publish on more frequent schedules: they used a quarterly schedule somewhat more often than open and mixed publishers, and also published much more often on monthly, bimonthly (interpreted here as every two months), and three times a year schedules than open and mixed publishers. Closed publishers were also far less likely to publish irregularly.

	Open publishers	Closed publishers	Mixed publishers
Continuous	3, 0.8%	1, 0.5%	7, 0.7%
Biweekly	1, 0.3%	1, 0.5%	
Weekly			1, 0.1%
Semi-monthly	1, 0.3%		1, 0.1%
Monthly	4, 1.1%	16, 7.5%	28, 2.7%
Bimonthly	13, 3.6%	20, 9.4%	55, 5.2%
Quarterly	73, 20.1%	57, 26.8%	229, 21.7%
Three times a year	20, 5.5%	18, 8.5%	62, 5.9%
Semiannual	118, 32.5%	41, 19.2%	295, 28%
Annual	54, 14.9%	21, 9.9%	175, 16.6%
Biennial	1, 0.3%	1, 0.5%	7, 0.7%
Other	20, 5.5%	9, 4.2%	45, 4.3%
Completely irregular	44, 12.1%	2, 0.9%	58, 5.5%
Not specified	11, 3%	26, 12.2%	91, 8.6%

Figure 22: The number and percentage of journals published by periodicity and publishers' OA status.

The distribution of the lifespans of open, closed, and mixed publishers are shown in Figure 23. Publishers' current ages were calculated by deducting the year at which a publisher's earliest journal was published from 2022. Closed publishers tended to be somewhat older than open and mixed publishers; the median age was 27 years and the majority of publishers were between 14 and 44 years old. Conversely, open publishers were typically younger with the majority aged between 9 and 24 years and a median of 15 years.

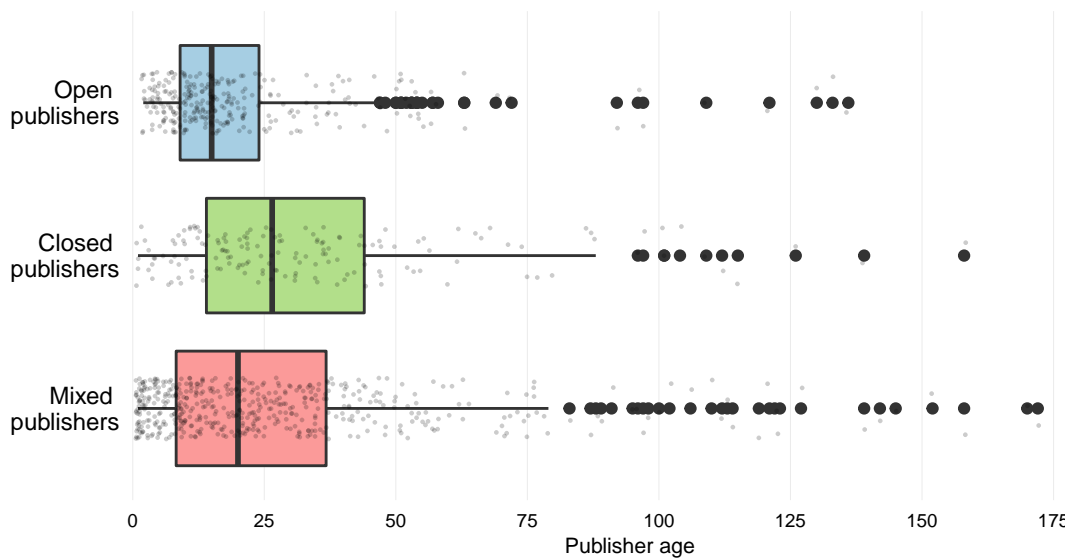


Figure 23: The age distribution of publishers of different OA statuses

Figure 24 shows the number and percentage of journals from publishers of different OA statuses that were ongoing, introduced, or discontinued during the 2019-2021 period. For all publishers the majority of journals were continued throughout these years and few journals (2-4%) were discontinued. However, mixed publishers introduced 17% of their journals in 2019-2021, far more than the 5-6% of open and closed publishers' journals introduced in these years, perhaps indicating stronger growth in this group.

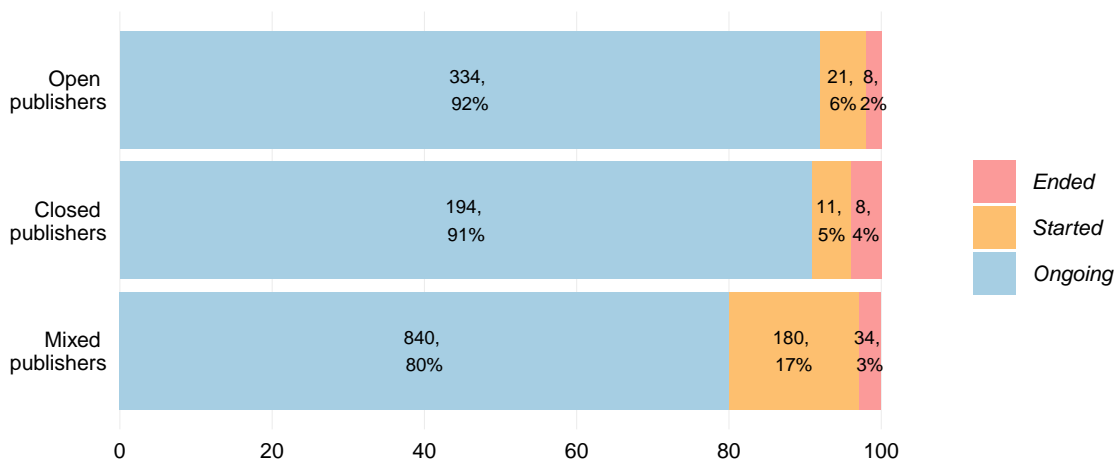


Figure 24: The number and percentage of journals started, ended, or continued through the 2019-2021 period by publishers' OA status.

The number and percentage of journals that changed publisher between the journals' first and latest publication events are shown in Figure 25, disaggregated by publishers' OA status. The patterns of journal ownership were quite similar between groups, with only 6-9% of journals changing publishers over time; the large majority of journals were retained by the same publisher. However, the status of 5-13% of journals was unknown.

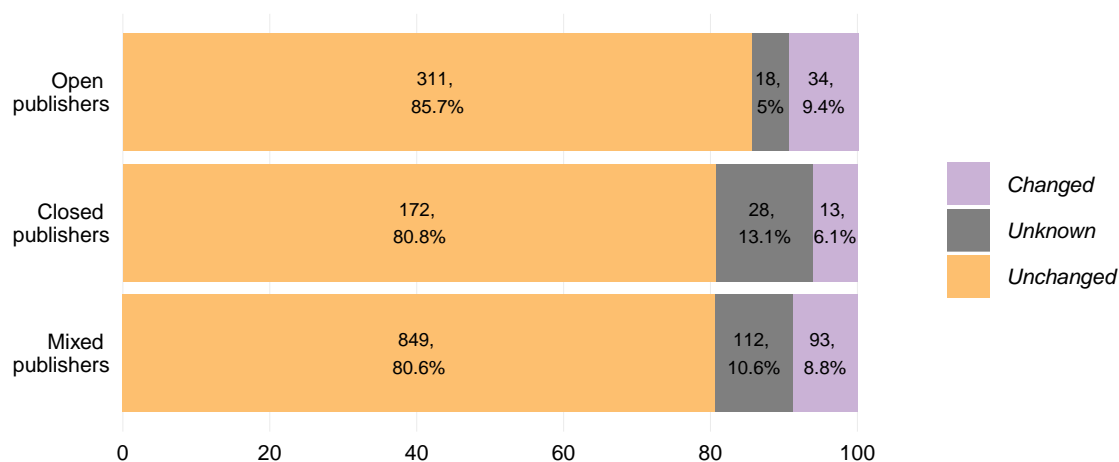


Figure 25: The number and percentage of journals and whether they changed publishers over time, by publisher OA status.

Conclusions and considerations

This Section of the study compared publishers of different OA statuses on several characteristics to generate a profile of publishers of each type. This information can assist the KE in observing differences between small publishers that have and have not begun a transition to OA, and identifying areas in which publishers may need particular support in transitioning to OA.

We assigned publishers an OA status based on the classification of their journals as open, closed or mixed using a threshold of the status of 90% of the journal's content in Unpaywall. Using this classification, we observed that the majority of publishers in most size categories and countries were mixed. The exception was Finland where two-thirds of publishers were open, and the Netherlands had nearly equal proportions of open and mixed publishers. Commonly, approximately half of countries' publishers were mixed, 20-30% were open, and a smaller percentage were closed publishers. Germany and France had higher percentages of closed publishers, however. Very small publishers also had similar levels of mixed and open publishers, compared to rather small and small publishers, which were predominantly mixed.

In comparing publishers of different OA status, differences emerged in their profiles of characteristics. For example, open and mixed publishers were more likely to publish material in English than were closed publishers, who published more often in local languages. Perhaps in contrast to this finding, open and mixed publishers had a somewhat stronger focus on humanities than did closed publishers. However, all publishers were clearly more oriented toward the social sciences and humanities than medical, technological, or natural science fields. Closed publishers were also typically older than mixed and particularly open publishers. Closed publishers also tended to publish their journals more frequently, such as on monthly, bimonthly, and quarterly schedules, compared to the open and mixed publishers that published two-thirds of their journals quarterly, semiannually, or annually. There were

fewer differences between publisher groups in terms of the continuation of journals and changes in publishers, with the majority of journals being continued and retained by the same publisher. However, mixed publishers appeared to be experiencing more growth, with substantially more of their journals introduced in the last three years than closed and open publishers' journals.

A consideration in the interpretation of these data, however, is the unevenness of the sample sizes for publishers of different OA types. The majority of publishers were mixed, and relatively few were closed. The effect of these differences is limited here as our analysis is descriptive, rather than statistical. However, interpretations of the results should consider that the sample of closed publishers is relatively small.

Overall conclusions

Overall, we conclude that Crossref represents a suitable database from which to identify a range of publishers in the KE countries. However, as discussed, incomplete registration of publishers' content in Crossref, the assignment of multiple memberships, publishers' use of alternate DOI registries, or lack of DOI registration entirely are all reasons why small publishers may not be identifiable in Crossref. As such, we recommend that future studies aiming to expand the coverage of small publishers examine alternate data sources that may include small publishers. This may include those sources identified in the first component of the study, such as Ulrichsweb or QOAM, or other DOI registries, such as DataCite. Nationally-oriented data sources, e.g. Mirabel³⁴ for France or journals.fi for Finland, may also be useful for coverage of specific countries. The utility of such lists was demonstrated in the current study as the supplementary lists of French and Dutch journals identified 8% of publishers, either by disambiguating the publisher from a much larger entity or identifying publishers not present in Crossref.

We also found Unpaywall to be an invaluable data source for assessing the OA status of documents, journals, and publishers. Its essentially complete coverage of publishers in the current study was influenced by the interdependencies between Crossref and Unpaywall. However, Unpaywall draws from over 50,000 data sources, suggesting its coverage may extend to other sources used in future studies, such as DataCite, and thus may continue to be a useful resource for OA information. The alignment between Unpaywall and the DOAJ suggests that the DOAJ captured the majority of OA journals and publishers in our sample. In contrast, small publishers were not particularly visible in the Sherpa Romeo database, especially French and German publishers. Further, the missingness of data in this source was often problematic for analyses. However, this issue was not specific to Sherpa Romeo. For instance, we observed the rate of missing license information in Crossref to be much higher in our sample of small publishers than in a sample examining publishers in Scopus and WoS[1]. As such, it may be the case that small publishers provide less metadata to repositories they are indexed in, which may hinder their identification and analysis of their characteristics.

Despite the incompleteness of data, distinct trends and profiles of small publishers of different sizes, countries, and OA statuses emerged. For instance, we observed that small publishers typically published their journals via a combination of OA pathways, with fewer publishing using only open or, in particular, closed models. This practice might reflect the transition of publishers to OA or represent a choice by the publisher to use multiple publishing pathways. Qualitative studies may be useful to distinguish between open or closed and these mixed publishers in their decision to publish along one or multiple pathways.

³⁴<https://reseau-mirabel.info/api>

It was common across publishers of all sizes and countries for their journals to be located primarily in the social sciences and humanities fields. This potential orientation toward more nationally or regionally relevant research questions was reflected in the prominent use of local languages by many publishers, far beyond the levels observed for publishers in WoS and Scopus. Indeed, French, Finnish, and closed publishers were more likely to publish in local languages than in English. Closed publishers also published their journals on more frequent schedules than did open and mixed publishers, who typically published journals 1 to 4 times a year. This may occur due to differences in costs between publishing models or the time required to collect sufficient content. However, qualitative studies could inform us of the decision-making of publishers in their varying approaches to these schedules.

Across indicators, it was notable that Finland regularly emerged as a positive exception, with a higher proportion of OA journals and publishers than other countries. As such, future qualitative studies may wish to explore the characteristics of Finnish publishers and journals to identify the features influencing their transition to OA. For instance, anecdotally, the early sign-on to PlanS by the Academy of Finland has positively influenced the OA transformation through efforts by Finnish publishers to attain or retain PlanS compliance, as has the introduction of the journals.fi platform and other local and funder policies. In contrast, German and French publishers had higher levels of closed publishers than other countries, which may warrant further examination to understand the factors underlying this trend.

Taken together, our findings highlight how crucial small publishers are to bibliodiversity, in particular by their strong contributions in the social sciences and humanities and the use of local languages, both of which are often overlooked and under-represented in the broader academic community. Small publishers thus ensure richness and diversity in academic discourse by providing critical communication channels for areas with specific cultural or linguistic themes. Consequently, it is imperative that small publishers are supported to continue their important contribution to academia as the community moves toward an open model of research.

References

- [1] A. Akbaritabar and S. Stahlschmidt. “Merits and limits: Applying open data to monitor open access publications in bibliometric databases”. In: *17th International Conference on Scientometrics and Informetrics (ISSI)*. Rome, Italy, Sept. 2019.
- [2] A. Becerril and et al. *OA Diamond Journals Study. Part 2: Recommendations*. en. Recommendations 2. cOALITION S, Mar. 2021, p. 36. DOI: 10.5281/ZENODO.4562789. URL: <https://zenodo.org/record/4562789> (visited on 09/06/2022).
- [3] J. Bosman et al. *The OA Diamond Journals study. Part 1: Findings*. <https://zenodo.org/badge/DOI/10.5281/zenodo.4553103.svg>. Retrieved 31 March 2022. 2021.
- [4] X. Chen. “Scholarly Journals’ Publication Frequency and Number of Articles in 2018–2019: A Study of SCI, SSCI, CSCD, and CSSCI Journals”. In: *Publications 7.3* (2019). DOI: 10.3390/publications7030058.
- [5] Crossref. *Fact File: 2018-2019 Annual Report*. <https://www.crossref.org/pdfs/annual-report-factfile-2018-19.pdf>. Retrieved 21 December 2021. 2019.
- [6] L. Estelle, D. Jago and A. Wise. *How to enable smaller independent publishers to participate in OA agreements*. <https://bit.ly/3xhd7kx>. Retrieved 21 December 2021. 2021.
- [7] A. Gazni. “Globalization of national journals: investigating the growth of international authorship”. In: *Learned Publishing 28* (2015), pp. 195–204. DOI: <https://doi.org/10.1087/20150305>.
- [8] N. Jahn. *roadoi: Find Free Versions of Scholarly Publications via Unpaywall*. R package version 0.7.2. 2022. URL: <https://CRAN.R-project.org/package=roadoi>.
- [9] M. Laakso and A-M. Multas. “European scholarly journals from small- and mid-size publishers in times of Open Access: Mapping journals and public funding mechanisms”. In: *Zenodo* (2022). DOI: 10.5281/zenodo.5909512.
- [10] V. Larivière, S. Haustein and P. Mongeon. “The oligopoly of academic publishers in the digital era”. In: *PLOS One 10.6* (2015), e0127502. DOI: 10.1371/journal.pone.0127502.
- [11] V. I. Levenshtein. “Binary codes capable of correcting deletions, insertions and reversals.” In: *Soviet Physics Doklady 10.8* (1966), pp. 707–710.
- [12] A. Pacher. “List of academic publishers and their scholarly journals: a webscraping approach”. In: *SocArXiv* (2021). DOI: 10.31235/osf.io/56b28.
- [13] N. Quaderi et al. *The Plan S footprint: Implications for the scholarly publishing landscape*. <https://beopen-project.eu/storage/files/ws190021-isi-report-2019-013.pdf>. Institute for Scientific Information. 2019.
- [14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020. URL: <https://www.R-project.org/>.
- [15] D. Robinson. *fuzzyjoin: Join Tables Together on Inexact Matching*. R package version 0.1.6. 2020. URL: <https://CRAN.R-project.org/package=fuzzyjoin>.
- [16] K. Shearer et al. *Fostering bibliodiversity in scholarly communications: A call for action!* <https://doi.org/10.5281/zenodo.3752923>. 2020.
- [17] D. Stephen, S. Stahlschmidt and S. Hinze. “Performance and Structures of the German Science System. Studien zum deutschen Innovationssystem (aktualisierte Version Mai 2020).” In: EFI, 2020. Chap. Studie 5-2020.

- [18] G. Toledo et al. *Bibliodiversity - what it is and why it is essential to creating situated knowledge*. <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/05/bibliodiversity-what-it-is-and-why-it-is-essential-to-creating-situated-knowledge/>. Impact of Social Sciences Blog. 2019.