Check for updates

The MIT Press

Crossref as a bibliographic discovery tool in the arts and humanities

Ángel Borrego[1] 🆔, Jordi Ardanuy[1] 🆔, and Llorenç Arguimbau[2] 🆔

[1]Departament de Biblioteconomia, Documentació i Comunicació Audiovisual & Centre de Recerca en Informació, Comunicació i Cultura (CRICC), Universitat de Barcelona, Barcelona, Spain
[2]ContextI+D, Cerdanyola del Vallès, Spain

ABSTRACT

Crossref is an official digital object identifier registration agency launched in 2000 as a joint effort between publishers to allow persistent cross-publisher citation linking in online academic journals. Our study explores the coverage of Crossref for tracking literature in the arts and humanities, which usually has a national or regional focus and targets domestic audiences. An analysis of the coverage of ERIH PLUS journals shows that Crossref indexes more sources than Scopus and includes additional journals from Eastern and Southern Europe and the Global South. Crossref limitations arise when analyzing the amount of metadata deposited by publishers. Just two-thirds of the journals deposit abstracts and ORCIDs and around a third deposit affiliations. The level of metadata completion for individual articles is lower, with major differences depending on the language of the document. Just half of the journals actually deposit references. As a result, Scopus retrieves more citations than Crossref, except for publications in German and French. Crossref represents a promising bibliographic discovery tool in the arts and humanities but is in need of improvement regarding the level of metadata completion.

1. INTRODUCTION

Several bibliographic data sources have appeared in recent years, thereby diversifying the set of tools available for searching for academic literature. In contrast to traditional bibliographic databases provided by commercial companies such as Scopus (Baas, Schotten et al., 2020) and Web of Science (Birkle, Pendlebury et al., 2020), some of these bibliographic information providers offer metadata available openly to the public. These metadata are license-free because metadata are facts: They cannot be owned, and therefore they have no license. One of the most important open metadata infrastructure systems in this information landscape is Crossref, an official digital object identifier (DOI) registration agency (Hendricks, Tkaczyk et al., 2020).

Crossref[1] is a not-for-profit association that provides most persistent identifiers assigned to academic publications and publishes the metadata associated with these publications. It was launched in 2000 as a collaborative effort by publishers to enable persistent cross-publisher citation linking between academic journals. DOIs are used to uniquely identify digital objects

---

[1] https://www.crossref.org.

(articles, data sets, monographs, reports, etc.). A DOI takes the form of a character string divided into two parts, a prefix and a suffix, separated by a slash (e.g., 10.1000/173). The DOI remains fixed over the lifetime of the document and is tied to its metadata, including the URL, thus providing access to the document. Referring to an online document by its DOI supposedly provides a more stable link than simply using its URL. However, for this to happen, publishers must update metadata in the event of a change in URL so that the DOI links to the new URL.

The open availability of document metadata through Crossref underlies the proliferation of new academic information services (Martín-Martín, 2021): academic search engines such as Dimensions and Lens; reference management software such as Zotero; and services to identify open access versions of academic publications such as Unpaywall.

Multidisciplinary bibliographic databases such as Scopus and Web of Science have traditionally been criticized for their limitations in terms of tracking research in the social sciences and humanities (Mongeon & Paul-Hus, 2016). Research in these fields frequently has a national or regional focus and targets domestic audiences. As a result, a considerable number of academic publications in these fields are published in national or regional journals outside the coverage of Scopus and Web of Science (Nederhof, 2006). Several studies have identified an overrepresentation of English language journals and English-speaking countries and an underrepresentation of documents from the arts, humanities, and social sciences in both Web of Science and Scopus, although the latter has much wider coverage (Mongeon & Paul-Hus, 2016; Vera-Baceta, Thelwall, & Kousha, 2019). A quick search shows that 93.1% of articles, reviews, and proceedings indexed in Scopus in 2020 were in English, while the figure for Web of Science was 96.5%. Google Scholar provides better coverage than Scopus and Web of Science but is limited in terms of usage, thereby reducing its usefulness for large-scale citation analyses (Martín-Martín, Orduna-Malea et al., 2018).

Our study aims to explore the coverage of Crossref for tracking literature in the arts and humanities by analyzing its coverage of journals in ERIH PLUS[2], an index containing bibliographic information on academic journals in the social sciences and humanities. ERIH stands for "European Reference Index for the Humanities." However, in 2014, the list was renamed ERIH PLUS, to indicate that it had been extended to include social science disciplines as well. Because it is hard to draw a precise line between humanities and social sciences, we have considered all journals listed in ERIH PLUS. The inclusion of social science journals in the list should be borne in mind when analyzing the results. Crossref coverage was also compared with that of Scopus, with a special focus on geographical differences in the indexing of journals by both sources. Additionally, the amount of metadata present in Crossref records was measured for a sample of articles in the arts and humanities published in 2020 in eight languages. Finally, the number of citations to this sample of articles retrieved by Crossref and Scopus was compared.

2. BACKGROUND

2.1. What Is Crossref?

Crossref was created as a neutral party among publishers to enable the exchange of links between article reference lists through DOIs. It was envisioned as a digital archive of journals, accessible free of charge and with the added value of reference linking (Crossref, 2009, p. 8).

---

[2] https://kanalregister.hkdir.no/publiseringskanaler/erihplus/.

The metadata deposited by publishers for bibliographic works includes the reference lists. Crossref uses these references to create links between works that cite each other. The number of citations each work receives is visible to anyone through Crossref public APIs. In addition, Crossref members who deposit references can retrieve the full list of citing works (not just the count), and can display them on their website[3]. At present, journal content represents the largest subset of Crossref content, given that it accounted for 73% of the 106 million records registered in 2019 (Hendricks et al., 2020).

Crossref asks members to deposit as much rich metadata as possible, including the list of references. Until recently, members could choose whether their references were "closed" (only used for the "cited-by" service, but not distributed through any public interface), "limited" (organizations that signed an agreement for a subscription-based service could access these references) or "open" (available to everyone through open APIs) (Hendricks et al., 2020, p. 425). However, this "reference distribution preference" was removed and, since 3 June 2022, all references in Crossref are treated as open metadata[4].

The Initiative for Open Citations (I4OC)[5] is an advocacy group that campaigns to encourage publishers to make references of their academic publications openly available. Based on these data, the OpenCitations Index of Crossref open DOI-to-DOI citations (COCI)[6] has been developed (Heibi, Peroni, & Shotton, 2019; Peroni & Shotton, 2020).

2.2. How Does Crossref Compare to Other Bibliographic Data Sources?

One of the first systematic studies to compare Crossref to other bibliographic databases was conducted by Harzing (2019), who concluded that it might serve as a good alternative to Scopus and Web of Science, although Google Scholar and Microsoft Academic[7] were the most comprehensive free sources of bibliographic information. In a subsequent study, Chudlarský and Dvořák (2020) studied whether Crossref could replace Web of Science for research evaluation purposes using the Czech Technical University in Prague as a case study. They observed that just 53.7% of Web of Science citation links were present in COCI.

Martín-Martín, Thelwall et al. (2021) compared the coverage of more than three million citations to a sample of highly cited documents in six data sources. They concluded that Google Scholar was the most comprehensive source, whereas COCI was the smallest, given that it retrieved just 28% of all citations. However, an update in September 2021 showed that COCI coverage had increased to cover up to 53% of citations (Martín-Martín, 2021).

Van Eck and Waltman (2021) focused on the amount of metadata provided by Crossref to measure the availability of six elements in Crossref: reference lists, abstracts, ORCIDs, author affiliations, funding information, and license information. They observed that coverage had improved with respect to previous measurements, although there were significant differences in the submission of metadata among publishers. A subsequent study by Visser, Van Eck, and Waltman (2021) compared Crossref with four multidisciplinary bibliographic data sources:

---

[3] https://www.crossref.org/services/cited-by/.

[4] https://www.crossref.org/blog/amendments-to-membership-terms-to-open-reference-distribution-and-include-uk-jurisdiction/.

[5] https://i4oc.org.

[6] https://opencitations.net/index/coci.

[7] Microsoft Academic was discontinued at the end of 2021: https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/.

Dimensions, Microsoft Academic, Scopus and Web of Science. In terms of size, Crossref covered 35 million documents published in the period 2008–2017, which was substantially more than Scopus and Web of Science. However, in terms of references, 58% of the citation links in Scopus could not be retrieved from Crossref. As mentioned above, some publishers deposited documents in Crossref without references, but others did not make them openly available.

In the health sciences, Liang, Mao et al. (2021) investigated the coverage and citation quality of five freely available data sources for 30 million PubMed documents. Dimensions was the most comprehensive data source, given that it provided references for 62.4% of the documents, whereas COCI covered 34.7%.

Beyond comparative studies on the coverage of different data sources, the Ministry of Education and Science of Ukraine launched the Open Ukrainian Citation Index (OUCI)[8], a search engine and citation database that comprises citations from all publishers that use Crossref's "cited-by" service (Cheberkus & Nazarovets, 2019). Based on this tool, Mryglod, Nazarovets, and Kozmenko (2021) conducted a disciplinary analysis of Ukrainian economic research based on Crossref data.

3. OBJECTIVES

This article aims to explore the coverage of academic publications in the arts and humanities in Crossref for tracking the literature in these fields, with a special focus on geographical and linguistic coverage. The study is underpinned by the following research questions:

1. To what extent are ERIH PLUS journals covered by Crossref?
2. How does the coverage of Crossref compare with that of Scopus for ERIH PLUS journals?
3. Are there any geographical differences in the coverage of both sources?
4. To what extent are the metadata of individual articles deposited in Crossref?
5. How does the number of citations received by articles in Crossref compare to the number of citations received in Scopus?

4. METHODS

4.1. Journal-Level Comparison

A possible approach for comparing the coverage of several bibliographic databases would be to record all the journals indexed by each source in a single list. It would then be possible to measure the extent to which each source covers the whole set of journals. This approach was not feasible in our study, as Crossref does not support subject searching[9] and it was therefore not possible to identify the arts and humanities journals indexed. Instead, we used ERIH PLUS as the initial source of journals and measured the coverage of Crossref and Scopus against this list.

ERIH PLUS is an index that holds bibliographic information on academic journals in the social sciences and humanities. Journals submitted for inclusion in ERIH PLUS are evaluated based on several criteria related to editorial quality, authorship, transparency, etc. At the time of data collection, February 2022, ERIH PLUS listed 10,213 journals.

---

[8] https://ouci.dntb.gov.ua/en/.
[9] https://community.crossref.org/t/retrieve-subjects-and-subject-from-journals-and-works/2403.

The set of journals listed in ERIH PLUS was compared with the journals indexed in Scopus considering print and online ISSNs[10]. Similarly, all ISSNs were searched in Crossref through its public API using the R package *rcrossref*[11].

To identify any geographical differences in the coverage of journals by both sources, we classified the journals' countries of publication provided by ERIH PLUS according to the geographical regions used by the United Nations Statistics Division in its publications and databases. This division compiles and disseminates global statistical information, develops standards and norms for statistical activities, and supports countries' efforts to strengthen their national statistical systems[12].

4.2. Article-Level Comparison

To determine the extent to which metadata (abstracts, ORCIDs, affiliations, funding, licenses, and references) were present in individual records, we built a sample of articles in the arts and humanities published in 2020. As Crossref does not support subject searching, we retrieved all journal content (mainly articles, but also reviews, editorials, letters, etc.) from Scopus with a DOI classified in the arts and humanities published in English in 2020 that had received three or more citations at the time of data collection in February 2022 ($n = 17,054$), and all journal content with a DOI classified in the arts and humanities in the seven languages with output of at least 1,000 documents in 2020: Spanish ($n = 7,330$), Russian ($n = 4,696$), French ($n = 3,330$), Italian ($n = 1,864$), Portuguese ($n = 1,760$), German ($n = 1,583$) and Polish ($n = 1,127$). The query used to retrieve the records from Scopus was as follows:

SUBJAREA(arts) AND DOI(10.*) AND (LIMIT-TO(SRCTYPE,"j")) AND (LIMIT-TO(PUBYEAR,2020))

We searched all DOIs in Crossref through its public API using the R package *rcrossref*. In addition, for the analysis of the metadata deposited, we compared the number of citations received by each article according to both sources, Scopus and Crossref. In the case of Crossref, we considered DOI to DOI citations, the ones recorded by the source. In the case of Scopus, we considered all citations, including the ones received from documents without a DOI.

4.3. Source Code and Data Availability

The R code to retrieve data from Crossref and to reproduce the analysis is available at https://github.com/angbor09/crossref_humanities/.

5. RESULTS

5.1. Journal Coverage

Scopus indexed 49% of the journals listed in ERIH PLUS, while Crossref indexed 80% (Table 1). There were, however, major differences in the coverage of journals by both sources depending on the world region in which journals are published. Thus, Scopus presented wide coverage of journals published in North America (86% of the journals listed in ERIH PLUS), Oceania (86%), Northern Europe (85%), and Western Europe (74%). These percentages were

---

[10] https://www.scopus.com/sources.uri.
[11] https://github.com/ropensci/rcrossref.
[12] https://unstats.un.org/UNSDWebsite/; https://unstats.un.org/unsd/methodology/m49/.

Table 1. ERIH PLUS journals covered by Scopus and Crossref by world region

World regions	ERIH PLUS journals	Journals in Scopus	% journals in Scopus	Journals in Crossref	% journals in Crossref
Africa	41	17	41	23	56
Asia	357	74	21	284	80
Europe (East)	2,280	547	24	1,666	73
Europe (North)	1,590	1,351	85	1,489	94
Europe (South)	1,843	778	42	1,210	66
Europe (West)	1,265	938	74	1,086	86
Latin America and the Caribbean	1,412	173	12	1,078	76
North America	1,231	1,064	86	1,154	94
Oceania	51	44	86	45	88
Not available	143	40	28	113	79
Total	**10,213**	**5,026**	**49**	**8,148**	**80**

lower for journals published in Southern Europe (42%), Africa (41%), Eastern Europe (24%), Asia (21%), and Latin America and the Caribbean (12%).

Crossref presented better coverage of journals published worldwide. Like Scopus, it covered ERIH PLUS journals published in North America (94%), Northern Europe (94%), Oceania (88%), and Western Europe (86%). Coverage was also wide for Asia (80%), Latin America and the Caribbean (76%), and Eastern Europe (73%). The regions with lowest coverage were Southern Europe (66%) and Africa (56%), although in both cases the coverage was higher than that provided by Scopus.

When detailing the metadata deposited by journal publishers, Crossref made a distinction between "backfile" records (i.e., those with a publication date older than 2 years) and "current" records (i.e., those published within the last 2 years)[13]. Table 2 details the metadata deposited by publishers of journals listed in ERIH PLUS for "current" records (i.e., for articles published in the past 2 years). When searching for an ISSN, Crossref returns a set of information for the journal, including logical fields ("True" or "False") indicating whether the journal deposits abstracts, ORCIDs, etc. The value is "True" as long as one article has an abstract (or an ORCID, etc.). For instance, the finding that 64% of journals deposit abstracts means that 64% of the journals had deposited at least one abstract.

The amount and type of metadata deposited in Crossref varied greatly depending on the world region in which the journal was published. Thus, journals published in Latin America and the Caribbean (86%), Southern Europe (83%), and Eastern Europe (75%) were most likely to deposit abstracts for their articles. Publishers in Northern Europe were most likely to deposit ORCIDs (78%) and affiliations (67%), whereas publishers in Latin American and the Caribbean tended to deposit ORCIDs (77%) but not affiliations (11%). The information on research

---

[13] https://github.com/CrossRef/rest-api-doc/issues/47.

Table 2. Metadata deposited in Crossref by publishers of journals listed in ERIH PLUS by world region

	Abstracts (%)	ORCIDs (%)	Affiliations (%)	Funders (%)	Licenses (%)	Open references (%)	References (%)
Africa	43	57	30	17	43	74	43
Asia	30	26	8	4	17	55	54
Europe (East)	75	58	24	3	38	59	34
Europe (North)	49	78	67	62	84	89	79
Europe (South)	83	67	12	8	44	54	30
Europe (West)	57	53	39	35	56	84	61
Latin America and the Caribbean	86	77	11	3	66	46	18
North America	46	59	48	51	59	76	66
Oceania	47	71	56	49	62	80	64
NA	69	67	26	18	50	68	47
Total	**64**	**64**	**33**	**26**	**56**	**68**	**49**

funding was most frequently deposited by journals published in Northern Europe (62%) and, to a lesser extent, in North America (51%). Publishers in Northern Europe (84%) usually deposited information on articles' licenses, whereas this information was provided to a lesser extent for journals published in other world regions.

Until June 2022, Crossref members could set reference distribution to open, limited, or closed. However, this setting was not linked to the actual submission of references. Most journals (68%), especially those in Northern Europe (89%) and Western Europe (84%), had used the default setting of open. However, just half of the journals (49%) actually registered references, whether open or not, for articles published in the past 2 years.

5.2. Article Metadata

The fact that a publisher has deposited metadata for articles published within the past 2 years does not mean that it has done so systematically for all its records. Therefore, to determine the extent to which publishers actually deposit metadata in Crossref, we built a sample of articles. Given the importance of domestic journals in the dissemination of academic publications in the arts and humanities and the inequalities observed in the coverage of journals in different world regions, we analyzed the presence of metadata for articles in English that had received three or more citations at the time of data collection and for all articles in the seven languages with output of more than 1,000 articles in Scopus in 2020 (Table 3).

Most of the arts and humanities articles indexed by Scopus in 2020 were also present in Crossref, with coverage ranging from 86% for articles in Polish to 99% for articles in English, which was the most frequent language in the sample. The only major exception was for articles in Italian, with just a quarter (27%) of the articles indexed in Scopus present in Crossref.

There were major differences in the amount and type of metadata deposited depending on the language of the document. Thus, most articles in Portuguese (81%), Spanish (71%), and

Table 3. Article metadata deposited by publishers in Crossref

Language	Articles in Scopus	Articles retrieved from Crossref	% articles retrieved from Crossref	Abstracts Crossref	% articles with abstract	Funding Crossref	% articles with funding info	License Crossref	% articles with license info	References Crossref	% articles with references
English	17,054	16,953	99	5,280	31	5,939	35	11,714	69	14,947	88
Spanish	7,330	6,698	91	4,735	71	76	1	2,844	42	1,503	22
Russian	4,696	4,430	94	1,944	44	41	1	397	9	286	9
French	3,330	3,031	91	498	16	26	1	580	19	928	21
Italian	1,864	510	27	255	50	1	0	114	22	103	20
Portuguese	1,760	1,697	96	1,369	81	8	0	1,274	75	653	45
German	1,583	1,453	92	649	45	68	5	244	17	238	14
Polish	1,127	965	86	651	67	2	0	329	34	69	7

Polish (68%) had an abstract, whereas these percentages dropped to 31% for articles in English. By contrast, 88% of the articles in English included references, the next highest rate being in Portuguese (45%). Thirty-five percent of the articles in English included funding information, but in other languages this information appeared only very rarely.

The presence or absence of metadata does not necessarily reflect the commitment of publishers to provide information. Some fields may not be applicable to certain articles. This is the case, for instance, with editorials or letters that lack abstracts or articles that do not acknowledge any source of funding. Therefore, figures in Table 3 cannot be assessed against the supposed ideal of 100% completion, although articles and reviews accounted for 96% of the documents in the sample. It is difficult to make comparisons with Scopus given its export limits. However, for a sample of the two thousand most cited articles in each language (*n* = 15,123), Scopus provided abstracts in 83% of the records and funding information in just 8%. As with Crossref, funding information in Scopus was mostly available for articles in English.

To determine the extent to which authors' ORCIDs and affiliations were deposited, we retrieved all authors from the sample of articles. We did not remove duplicates, but considered the presence of this information in the metadata of each article published by any given author. Table 4 shows significant differences in the presence of this information according to the language of the document. Thus, authors' ORCIDs were mostly present for outputs in Portuguese, Polish and, to a much lesser extent, Spanish. By contrast, affiliations were present in German and English publications, although in neither case did this information reach half of the authors. Scopus included affiliation metadata for 83% of the records in a sample of the most cited articles in each language (*n* = 15,123).

5.3. Number of Citations

Finally, we compared the number of citations received by each article according to both sources, Scopus and Crossref (Table 5). To make the comparison meaningful, we restricted the analysis to articles present in both sources.

Most of the articles in the sample were in English and all of them had received at least three citations at the time of data collection. For outputs in this language, Scopus presented a

**Table 4.**   ORCID and affiliation metadata deposited by publishers in Crossref

Language	Authors	ORCIDs	% ORCIDs	Affiliations	% affiliations
English	51,835	13,458	26	21,969	42
Spanish	9,786	3,651	37	732	7
Russian	9,057	1,362	15	808	9
French	4,706	187	4	301	6
Italian	618	44	7	133	22
Portuguese	2,602	1,801	69	916	35
German	1,705	37	2	735	43
Polish	1,065	666	63	30	3

Table 5. Citations to articles in Scopus and Crossref

Language	Articles	Citations in Scopus	Citations in Crossref
English	16,953	124,208	117,286
Spanish	6,698	1,708	959
Russian	4,430	1,150	582
French	3,031	1,145	1,169
Italian	510	69	32
Portuguese	1,697	260	205
German	1,453	296	337
Polish	965	103	76

minor advantage over Crossref, given that it retrieved 6% more citations (Figure 1). For outputs in other languages, there was no clear pattern. Crossref retrieved more citations than Scopus for documents in German (+14%) and French (+2%), whereas Scopus retrieved more citations for the remaining languages. In Russian (+98%) and Spanish (+78%), Scopus had nearly double the number of citations retrieved by Crossref. In the remaining languages, the number of outputs and citations was very small, thus limiting the meaningfulness of the results.

We compared the number of citations received by each output according to both sources, although we did not analyze overlaps in citations in the two databases. Nevertheless, Figure 2 shows a high level of association between the number of citations received by each output in both sources for articles in English and French. The relationship was much weaker for articles in other languages.
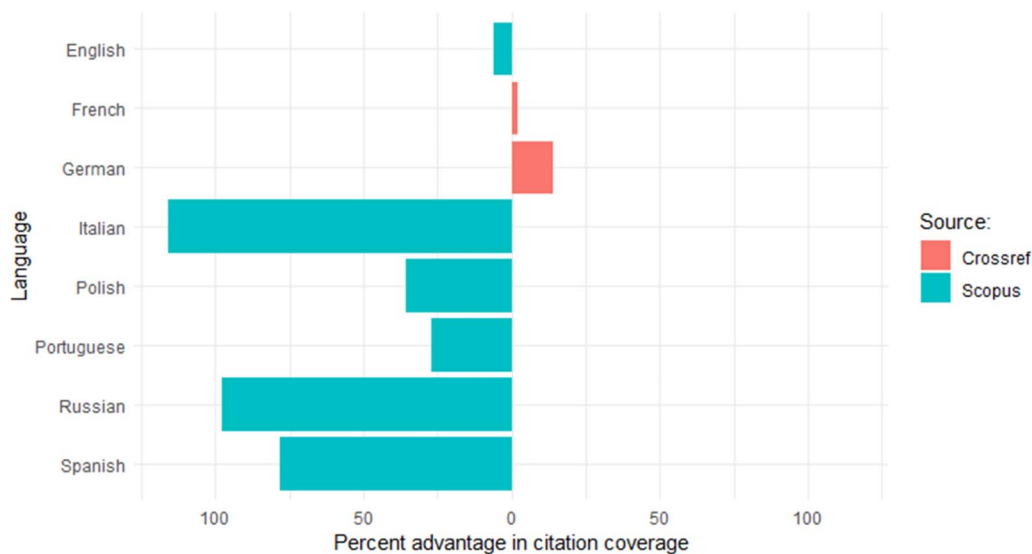
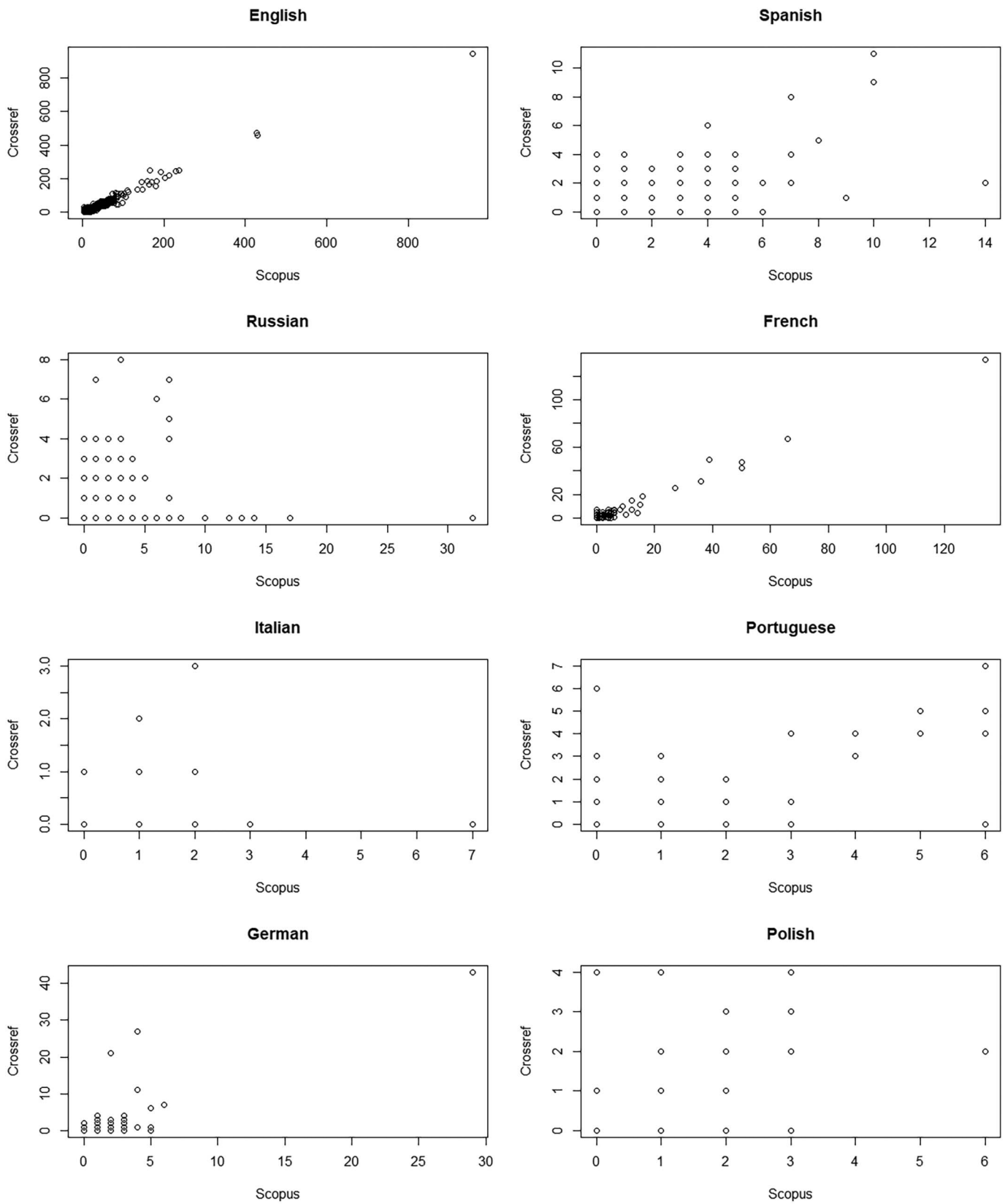**Figure 1.** Citations by language of the publication.

Figure 2. Citations received by each output according to Scopus and Crossref by language of the publication.

6. DISCUSSION AND CONCLUSIONS

The results of our study illustrate the advantages and limitations of Crossref as a source of bibliographic information in the arts and humanities. Crossref is an open resource built on the information deposited by publishers. It indexes a larger share of ERIH PLUS journals than Scopus. The additional journals covered are published mainly in Eastern and Southern Europe and the so-called Global South (i.e., Africa, Asia, and Latin America)[14]. These results are consistent with those of previous studies (Mongeon & Paul-Hus, 2016; Vera-Baceta et al., 2019) that have revealed an overrepresentation of English language journals in Scopus, and are noteworthy given that research in the arts and humanities frequently has a national or regional focus and is published in domestic journals.

When searching for individual articles, the overwhelming majority of those indexed in Scopus were also available in Crossref. The only major exception was articles in Italian, which presented a very low level of coverage in Crossref. A series of online searches suggest that Italian scholarly journals may be registering their DOIs not with Crossref but with another DOI registration agency, namely mEDRA, "a brand of ediSer, the service company of the Italian Publishers Association"[15].

The limitations of Crossref became evident when we analyzed the amount of metadata actually deposited by publishers. Less than two-thirds of the journals were found to deposit abstracts, and those that did deposit this information did not do so systematically for all articles. Slightly more than half deposited license information, which is relevant to measure compliance with open access mandates and open access availability.

The situation was similar regarding author information. Around two-thirds of the journals deposited ORCIDs and a third deposited affiliations. However, the level of metadata completion for individual articles was much lower, with major differences depending on the language of the document.

The inclusion of reference lists in records is important to improve retrieval options and for citation analysis. Our results suggest that most publishers were willing to share this information and make the reference lists in their journals openly available, although they could opt to make them "closed" or "limited." Nevertheless, this situation has changed recently, because new Crossref policies oblige publishers to make their references open. However, only half the journals actually deposit lists of cited references in their articles.

Although it could be surmised that the significant presence of journals published outside the Anglosphere[16] in Crossref could increase the amount of citation data for outputs in non-English languages, our results suggest that this is not the case. Except for outputs in German and French, Scopus retrieves more citations than Crossref, possibly because most publishers do not deposit reference lists. When interpreting this information, it should be borne in mind that Crossref only considers DOI to DOI citations, whereas Scopus also includes references received from documents without a DOI.

In summary, Crossref represents a promising source of information but is in need of improvement as a bibliographic discovery tool in the arts and humanities. The number of journals and articles indexed is vast and includes a large share of journals published outside North America and Europe. However, the amount of metadata deposited by publishers remains

---

[14] https://en.wikipedia.org/wiki/Global_North_and_Global_South.
[15] https://www.medra.org.
[16] https://en.wikipedia.org/wiki/Anglosphere.

limited. Further research could examine the motivations behind publishers' behavior in order to make Crossref a more comprehensive, accurate, and up-to-date source of information.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Ángel Borrego: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing. Jordi Ardanuy: Funding acquisition, Writing—review & editing. Llorenç Arguimbau: Funding acquisition, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

The data and code used in this study are available at https://github.com/angbor09/crossref _humanities/.

REFERENCES

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a _00019

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–376. https:// doi.org/10.1162/qss_a_00018

Cheberkus, D., & Nazarovets, S. (2019). Ukrainian open index maps local citations. *Nature*, *575*(7784), 596–596. https://doi .org/10.1038/d41586-019-03662-6, PubMed: 31772366

Chudlarský, T., & Dvořák, J. (2020). Can Crossref citations replace Web of Science for research evaluation? The share of open citations. *Journal of Data and Information Science*, *5*(4), 35–42. https://doi.org/10.2478/jdis-2020-0037

Crossref. (2009). The formation of CrossRef: A short history. https:// www.doi.org/topics/CrossRef10Years.pdf

Harzing, A. W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, *120*(1), 341–349. https://doi.org/10.1007/s11192-019 -03114-y

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations.

Scientometrics, *121*(2), 1213–1228. https://doi.org/10.1007 /s11192-019-03217-6

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414–427. https://doi.org/10 .1162/qss_a_00022

Liang, Z., Mao, J., Lu, K., & Li, G. (2021). Finding citations for PubMed: A large-scale comparison between five freely available bibliographic data sources. *Scientometrics*, *126*(12), 9519–9542. https://doi.org/10.1007/s11192-021-04191-8, PubMed: 34720252

Martín-Martín, A. (2021). La cobertura de los índices de citas abiertos se acerca a la de Web of Science y Scopus. *Anuario ThinkEPI*, *15*. https://doi.org/10.3145/thinkepi.2021.e15e04

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10 .1016/j.joi.2018.09.002

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*(1), 871–906. https://doi.org/10.1007/s11192-020 -03690-4, PubMed: 32981987

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Mryglod, O., Nazarovets, S., & Kozmenko, S. (2021). Universal and specific features of Ukrainian economic research: Publication analysis based on Crossref data. *Scientometrics*, *126*(9), 8187–8203. https://doi.org/10.1007/s11192-021-04079-7

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, *1*(1), 428–444. https://doi.org/10.1162/qss_a_00023

Van Eck, N. J., & Waltman, L. (2021). Crossref as a source of open bibliographic metadata. In *Proceedings of the 18th International Conference of the International Society for Scientometrics and Informetrics* (pp. 1169–1174). https://www.issi-society.org/proceedings/issi_2021/Proceedings%20ISSI%202021.pdf#page=1201

Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813. https://doi.org/10.1007/s11192-019-03264-z

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112