The MIT Press

RESEARCH ARTICLE

The availability and completeness of open funder metadata: Case study for publications funded by the Dutch Research Council

Bianca Kramer[1] iD and Hans de Jonge[2] iD

[1]Utrecht University Library, Utrecht, Netherlands
[2]Dutch Research Council, The Hague, Netherlands

ABSTRACT

Research funders spend considerable efforts collecting information on the outcomes of the research they fund. To help funders track publication output associated with their funding, Crossref initiated FundRef in 2013, enabling publishers to register funding information using persistent identifiers. However, it is hard to assess the coverage of funder metadata because it is unknown how many articles are the result of funded research and should therefore include funder metadata. In this paper we looked at 5,004 publications reported by researchers to be the result of funding by a specific funding agency: the Dutch Research Council NWO. Only 67% of these articles contain funding information in Crossref, with a subset acknowledging NWO as funder name and/or Funder IDs linked to NWO (53% and 45%, respectively). Web of Science (WoS), Scopus, and Dimensions are all able to infer additional funding information from funding statements in the full text of the articles. Funding information in Lens largely corresponds to that in Crossref, with some additional funding information likely taken from PubMed. We observe interesting differences between publishers in the coverage and completeness of funding metadata in Crossref compared to proprietary databases, highlighting the potential to increase the quality of open metadata on funding.

1. INTRODUCTION

Research funders spend considerable efforts on collecting information about the outcomes of the research they fund. Data about publications are among important data they collect because these represent direct results of research funding. These data can serve multiple purposes. Accountability is an important one, as governments increasingly expect funders to account for the impact of their funding and the efficiency of their operations. But publication data can also be an important element to inform strategy development. With the rise of open access, the collection of publication data became important for funders who want to track progress or check compliance with their open access policies.

The collection of publication data, especially how to link publications to grants, is associated with all kinds of complexity (Mugabushaka, 2020). Most, if not all, funding agencies require their grant holders to report publications associated with their funding. Many have invested in dedicated applications to capture these outputs, such as the RePORTER database of the National Institute of Health (NIH) or Researchfish, an application used by many UK funding councils. Reporting output through these systems is considered a burden by many

researchers because very often they are required to report the very same information at their home institution. In addition, there are concerns around the role of commercial players, such as Researchfish, in collecting these data (Inge, 2022). There have been initiatives to integrate current research information systems (CRIS) of universities with funder systems, but they do not seem to have been very successful yet (Clements, Reddick et al., 2017).

Consequently, many funders cannot guarantee they have a full picture of the outputs arising from their funding. Partly in response to that, commercial bibliographic databases have started to invest in providing information about links between funding and outputs based on the funding acknowledgment paratext of articles. In 2008, Web of Science (WoS) was the first to start collecting funding text, funding organization, and grant numbers systematically. Five years later it was followed by Scopus (Álvarez-Bornstein & Montesi, 2021). More recently, Digital Science launched Dimensions, the paid version of which seems to be explicitly developed for the market of funding organizations seeking insight into the outputs of their grants. Dimensions aims to provide information on the connections between publications, awarded grants, data sets, and other outputs from the larger research life cycle (Herzog, Hook, & Konkiel, 2020). There is a large body of literature trying to assess the completeness of these databases (Álvarez-Bornstein & Montesi, 2021; Grassano, Rotolo et al., 2017; Liu, Tang, & Hu, 2020).

With the launch of Crossref's service to collect and share funding information, an interesting new—open—data source of funding information became available (Lammey, 2014; Meddings, 2013). As from 2013 it is possible for publishers to add funding information to the standard Crossref metadata when registering a DOI, or when updating metadata for existing records. These funding data can be obtained from authors when they submit a manuscript or extracted from acknowledgment sections of manuscripts. Publishers are expected to provide information for three elements: "funder_name," "funder_identifier," and "award_number." Funding information is expected to be submitted to Crossref in XML format as part of the initial metadata deposit, or added later in CSV format as supplemental metadata upload.

The share of Crossref records with funding data has steadily increased to reach about 25% in 2021[1], making it an increasingly interesting source for bibliographic metadata, especially because it is open, meaning that the data are freely available for anyone to use and reuse, and do not require a paid license (as is the case with the commercial providers mentioned above). One example of using these metadata is Lens, a bibliographic database that makes use of open bibliographic metadata (including funder information) from Crossref and other sources.

Little is known, however, about the completeness of the funding data in Crossref. Clearly a lot of progress has been made over the past few years (Habermann, 2019). However, it is hard to evaluate the current rate of 25% for 2021 as we do not have a baseline against which to compare that figure. Not all papers registered in Crossref will be the result of external funding and therefore are not expected to contain funding information[2]. The percentage of records for which funding metadata are available will therefore never reach 100%.

In a recent paper, van Eck and Waltman (2021) have shown that there are large variations in the availability of funding information between publishers. A group of larger society presses (American Chemical Society (ACS), American Physical Society (APS), Royal Society of

---

[1] http://api.crossref.org/types/journal-article/works?rows=0&filter=from-pub-date:2021-01,until-pub-date:2021-12,has-funder:true.

[2] Of course, even "unfunded" research is often somehow funded in some way. However, convention has it that usually only external funding is acknowledged in acknowledgment sections, and not the (intramural) funding support from the authors' employers (Grassano et al., 2017).

Chemistry (RSC), and Optical Society) make funding information available for nearly all their articles. But there also seems to be a large group of smaller publishers that do not provide funding information at all. The larger publishers (Elsevier, Springer-Nature, Wiley, Taylor & Francis) attain percentages of around 40–50%. These observations are confirmed in a recent paper in which the coverage of funder information in Crossref is studied on the basis of the CORD-19 data set, a collection of publications and preprints on Covid-19 (Mugabushaka, van Eck, & Waltman, 2022).

This paper takes another approach to assess the completeness of open funder data in Crossref. As a basis we take a set of papers reported by researchers to be the result of external funding and therefore—in theory—should all contain funder metadata. The set of articles we use are those that were reported in 2021 by grant holders as resulting from funding by the Dutch Research Council, NWO, the major funding council of the Netherlands.

We will see that a majority of publications do contain funder metadata in Crossref, but still a substantial share of records do not. In addition, not all publications with funder metadata identify NWO as a funder either by including the organization's name and acronym or by using NWO's funder ID in the metadata. We also observe interesting differences between publishers. We have also compared the availability of funder data in Crossref with the major bibliographic data sources: Scopus, WoS, Dimensions, and Lens.

The importance of this analysis is that for the first time it provides us with a baseline. There is no doubt that the open availability of funder information has increased substantially in recent years. However, we did not know how good the coverage of funder metadata in Crossref is. Previous studies (Mugabushaka et al., 2022; van Eck & Waltman, 2021) generally take as the denominator all publications in a given data set, irrespective of whether they could or should contain funder information[3]. As argued above, not all publications can be expected to contain funder information as they are not externally funded or are publication types other than research papers (e.g., letters or editorials) for which acknowledging funding is unusual.

In contrast, the data set used for this analysis should—in theory—all have contained open funder data, because all publications have been reported by grantees as the result of funding by NWO. If our analysis is representative for all articles resulting from external funding in Crossref, it points to a sizable proportion of records that lack funding information where it could and should have been provided.

2. DATA

2.1. Data Set of Publications Resulting from NWO Funding

For this analysis we made use of a data set containing all peer-reviewed articles registered with the NWO in the year 2021. Like most funders, NWO requires the use of a funding acknowledgment in every publication[4]. All grantees are expected to register publications arising from their project using the grant management system ISAAC. In 2021, 5,530 publications were registered in the category of peer-reviewed journal articles[5]. Of these 5,530 articles, 157 did not have a DOI and were therefore left out of this analysis.

---

[3] An alternative approach to tackling the "denominator problem" is offered by Mugabushaka et al. (2022), where the authors have manually compared funding metadata with funding statements in the full text of papers.

[4] https://www.nwo.nl/en/acknowledgement-publications.

[5] NWO expects its grantees to report all outputs of funded projects. ISAAC therefore allows for the registration of multiple publication types, including data sets. For this research we have only used publications registered as peer-reviewed articles.
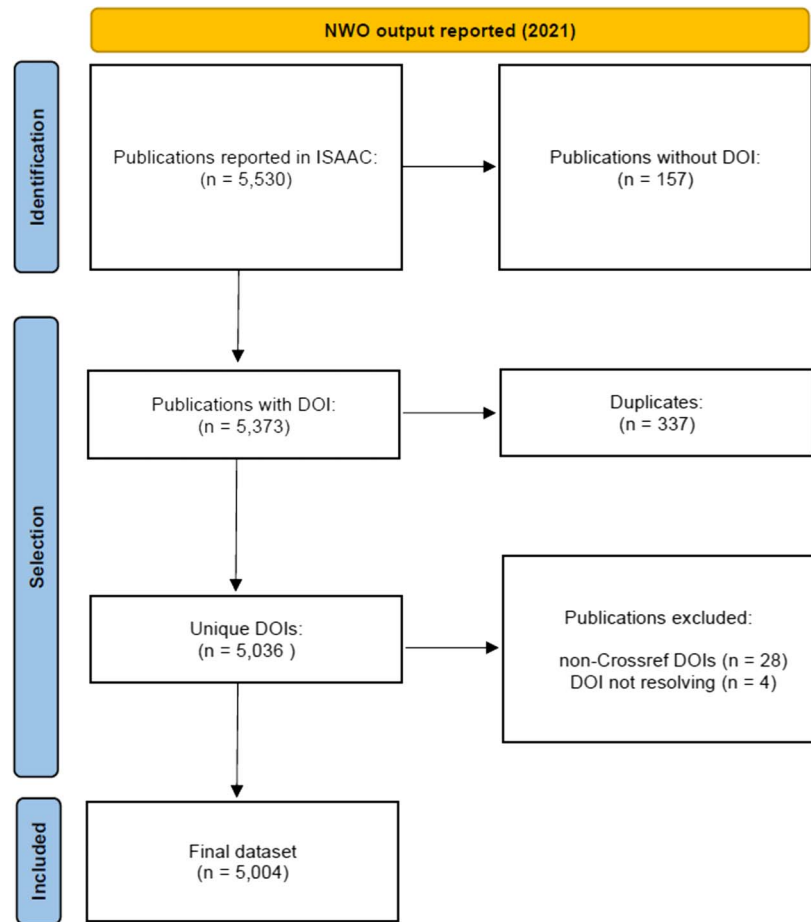
Figure 1. Composition of the data set of NWO-funded DOIs used for this analysis.

DOIs were cleaned using an R-script (see Data Availability), stripping url-prefixes (http://doi.org/ and https://dx.doi.org/), trailing punctuation, and additional text strings entered together with the DOI. After deduplication (because some publications were reported as outputs of multiple projects), 5,036 unique DOIs remained. Of these, 28 DOIs were issued by DOI registrars other than Crossref (such as DataCite and mEDRA), and four DOIs did not resolve, despite being listed as such on publishers' websites. The remaining 5,004 resolving Crossref DOIs made up the data set used in this study—see Figure 1.

2.2. Representativeness

In terms of disciplinary distribution, we cannot claim full representativeness. As a research council, NWO covers all disciplines. Medicine and health sciences, however, are funded by the Netherlands Organization for Health Research and Development, ZonMw, which is partly funded through NWO but has a separate system to capture the outputs of its funding. NWO itself is organized along three disciplinary domains: natural sciences, social sciences and humanities, and technical and engineering sciences. Table 1 shows the NWO domain of the funded projects that the publications in our data set resulted from. Around half the publications stem from the science domain, 37% from the social sciences and humanities, and a minority of 7% from technical and engineering sciences.

Table 1. Number and percentage of publications according to NWO domain

NWO domain	No.	%
Natural sciences	2,302	46.8
Multidisciplinary	370	7.4
Social sciences and humanities	1,846	36.9
Technical and engineering sciences	354	7.1
Unknown	94	1.9
Total	**5,004**	100.0

2.3. Retrieval of Metadata Including Funding Information

2.3.1. Crossref

For all records in the data set, metadata including funding information were retrieved using the Crossref REST API. The API was queried through a Google Apps script (using server side Java-script) (see Data Availability), returning the results directly to a Google Sheets document for further processing.

Metadata retrieved for each publication included publication type and year of earliest publication (online or in print), as well as member ID and the publisher's name associated with it. For each publication, the number of funders associated with it was retrieved, as well as all individual funder names and the presence or absence of one or more of the following funder IDs associated with NWO:

- 10.13039/501100003246 (Nederlandse Organisatie voor Wetenschappelijk Onderzoek)
- 10.13039/501100003958 (Stichting voor de Technische Wetenschappen)
- 10.13039/501100010409 (Nationaal Regieorgaan Praktijkgericht Onderzoek SIA)
- 10.13039/501100010071 (Nationaal Regieorgaan Onderwijsonderzoek)

To identify records with funding attributed to NWO in funder names, we manually identified all unique NWO-associated funder names present in the free-text "funder name" field in Crossref metadata for our data set. This included all primary and alternative labels for the funder IDs listed above, other variants of the full and abbreviated funder name in both English and Dutch, and the names of NWO's subdivisions in both English and Dutch, as well as the names of funding instruments unique to NWO. A complete list of identified funder name variants is available (see Data Availability). Funder names retrieved for each record were matched against this list and the presence or absence of one or more NWO-associated funder names recorded.

2.3.2. Lens

The list of 5,004 unique Crossref records was imported into a Collection in Lens (https://www .lens.org/)[6]. Collections allow batch import of up to 10,000 DOIs at once and batch export of selected metadata fields for up to 50,000. Metadata fields included in the export were Lens ID (unique identifier specific to Lens), DOI, and Funding. Funding information in Lens consists of a list of identified funders per record, from which the number of funders per record was

---

[6] The collection can be accessed using the following link: https://www.lens.org/lens/search/scholar/list?collectionId=200429.

calculated. No attempt was made at this time to specifically identify mentions of NWO in funding information from Lens.

2.3.3. WoS

Using Utrecht University's licensed instance of WoS Core Collection (containing the SCIE, SCI, AHCI, and ESCI citation indexes) was searched for the 5,004 unique Crossref DOIs. This was done by constructing a query for 1,000 DOIs at once, using the format DO=(10.1016/j .physletb.2020.135632 OR 10.4000/crcv.18857). The results were exported as a tab-delimited file containing the field Funding Information. Funding information in the database export contains the fields FU (harmonized funder names with grant numbers where available) and FX (free text funding acknowledgment). From the FU field, the number of funders per record was calculated. No attempt was made at this time to specifically identify mentions of NWO in the FU or FX fields.

2.3.4. Scopus

Using Utrecht University's licensed instance of Scopus, the database was searched for the 5,004 unique Crossref DOIs. This was done by constructing a query for 1,000 DOIs at once, using the format DOI(10.10.1016/j.physletb.2020.135632) OR DOI(10.4000/crcv.18857). Results were exported as a CSV file containing the fields categorized under Funder details (Number, Acronym, Sponsor, Funding text). Funding information in the database export contains the fields Funding Details (harmonized funder names with grant numbers where available) and Funder Text (free text funding acknowledgment). From the Funding Details field, the number of funders per record was calculated. No attempt was made at this time to specifically identify mentions of NWO in the Funding Details or Funder Text field.

2.3.5. Dimensions

The authors received temporary access to the licensed instance of Dimensions, through its No Cost Access program. The database was queried for all 5,004 papers in our data set in batches of 400 through the Dimensions API Connector, a Google Sheets add-on, using the query "search publications where doi in [{range}] return publications[doi+funders] limit 400", with "range" denoting the cell range containing the DOIs to be queried. Returned funder information in JSON format was extracted and processed using a Google Apps script (see Data Availability). Funder information in Dimensions consists of a list of identified funders per record (using harmonized funder names). From this, the number of funders per record was calculated.

Data were collected between during 2022 for Crossref, Lens, WoS, and Scopus, and May 2022 for Dimensions.

3. RESULTS

3.1. Retrieval of DOIs and Funder Metadata

Figure 2 provides an overview of the number of DOIs from our data set that were retrieved from each database, as well as the number of DOIs with funding information in each database. Both Lens and Dimensions have virtually 100% coverage of DOIs in our data set, while coverage in WoS and Scopus is slightly lower at 92% and 94%, respectively. The presence of funding information in Crossref and the other bibliographic databases will be discussed in more detail in the sections below.
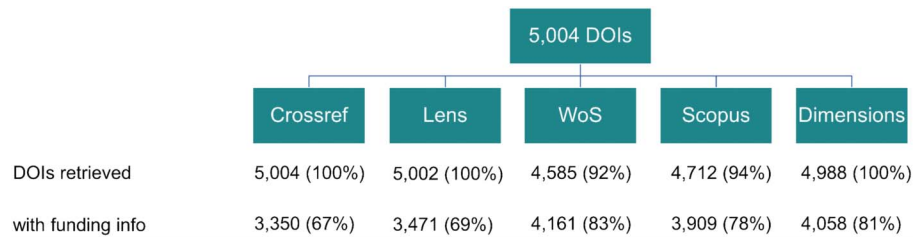
Figure 2. Retrieval of DOIs and funding metadata for the 5,004 DOIs in our data set for all databases studied.

3.2. Availability of Open Funder Data in Crossref

There is no doubt that the availability of funding information in Crossref has increased considerably since 2013, the year Crossref opened the possibility to register this information for its member organizations. Figure 3 presents an overview of the overall availability of funding information for our data set. Because grant holders do not necessarily register their publications in the same year a paper is published, our set contains articles published from multiple years (from 2011 to 2022). The majority of publications reported ($n = 3,660$, 73%) were published in 2020 and 2021[7].

Overall, 67% of the publications registered in 2021 contain some kind of funding information. This percentage is stable for articles published from 2018 onwards.
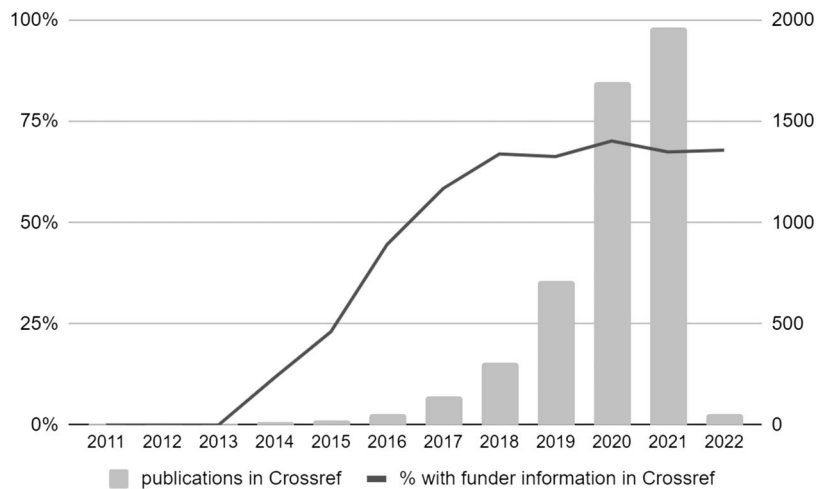

Figure 3. Crossref records in data set ($n = 5,004$), as well as percentage of DOIs that have funding information in Crossref, by publication year.

3.2.1. Presence of NWO funder name and NWO funder ID

Although 67% of the records in our data set contain funder information and these publications have been registered by NWO grantees as work funded by NWO, it often happens that among the funding organizations mentioned, NWO is not included. Only in 53% of the cases was NWO identified as a funder using the name of the organization (Figure 4). Although NWO has specific requirements for how the organization should be referred to when acknowledging funding, we detected no fewer than 174 name variations, including uppercase/lowercase variants.

---

[7] The data set predominantly consists of journal articles ($n = 4,906$, 98%), with a small number of preprints ($n = 55$), proceedings articles ($n = 14$), book chapters ($n = 12$) and other publication types ($n = 17$).
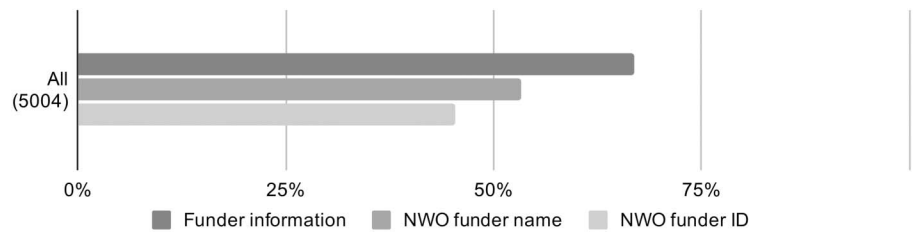
Figure 4. Crossref records in data set (*n* = 5,004) with funding information, NWO funder name and NWO funder ID in Crossref metadata.

Precisely to combat this enormous ambiguity in the names of funding organizations, Crossref has set up the Funder Registry. In this register all grant giving organizations are identified with a funder ID: a DOI for every single funding organization. The registry is an open data source, created and maintained by Elsevier, and can for instance be integrated in submission systems of publishers to allow authors to simply choose from a standardized list of funders. When registering funding information, Crossref expects publishers not only to provide the name of the funding organization but also its funder ID. There seems to be considerable room for improvement in this area: Only 45% of publications in our set was correctly attributed to NWO with the use of its funder ID. The majority of these IDs (94%) were asserted by the publisher and 6% by Crossref, which, as part of its data cleaning efforts, also tries to match Funder IDs with funder names where these are not provided by the publisher. In our data set, in 24% of the cases where a publisher did not provide the funder ID but did include a variant of NWO as funder name, Crossref was able to retroactively add the funder ID.

3.2.2. Differences by domain

It is well known from the literature (Álvarez-Bornstein & Montesi, 2021; Costas & Yegros, 2013; Grassano et al., 2017) that scientific fields have quite different cultures when it comes to acknowledging funding. It has been reported that publications from the social sciences and humanities are less likely to acknowledge external funding, possibly due to a lower availability of external funding compared to other fields. In contrast, publications from the natural sciences and medical sciences are more likely to acknowledge external funding because of the increased importance of being transparent about possible conflicts of interest with external funding of medical and health-related research. Interestingly, our data set does not show any major disciplinary differences. Table 2 shows that in the presence or absence of funding

Table 2. Number and percentage of DOIs that have funding information in Crossref according to NWO domain

NWO domain	No.	%
Natural sciences	1,626	69.5
Multidisciplinary	241	65.1
Social sciences and humanities	1,148	62.2
Technical and engineering sciences	284	80.2
Unknown	51	54.3
Total	**3,350**	**66.9%**

information in Crossref there are no major differences between publications from different NWO domains.

3.2.3. Differences per publisher

We also looked at how the different publishers perform when registering funding information with Crossref. As has been shown in earlier studies (Mugabushaka et al., 2022; van Eck & Waltman, 2021) there is a large variation in the degree to which publishers provide funding information to Crossref. Figure 5 provides an overview of the performance of the 20 biggest publishers in our data set, representing 86% of the DOIs in our data set.

Some of the larger society presses—ACS, APS, RSC—perform exceptionally well, with almost 100% of publications containing funding information. It is interesting to note that these publishers not only correctly identify NWO as the funder to a very high degree but also very often do so by using NWO's funder ID.

Next in line are the "big five" publishers: Elsevier, Springer-Nature, Wiley, Taylor & Francis, and SAGE, which provide some funding information for around 75% of their papers but perform considerably less well in correctly identifying NWO as the funder either by including the agency's name or the funder ID.

The larger full open access publishers (PLOS, MDPI, and Frontiers) seem to perform not much better when compared to the large legacy publishers: Around 75% of their DOIs contain funding information. But again, only in around 50% of the cases is NWO correctly identified as the funder of the research. This may come as a surprise, given the greater financial dependency of full gold open access publishers on funders such as NWO for APC payment and—consequently—on
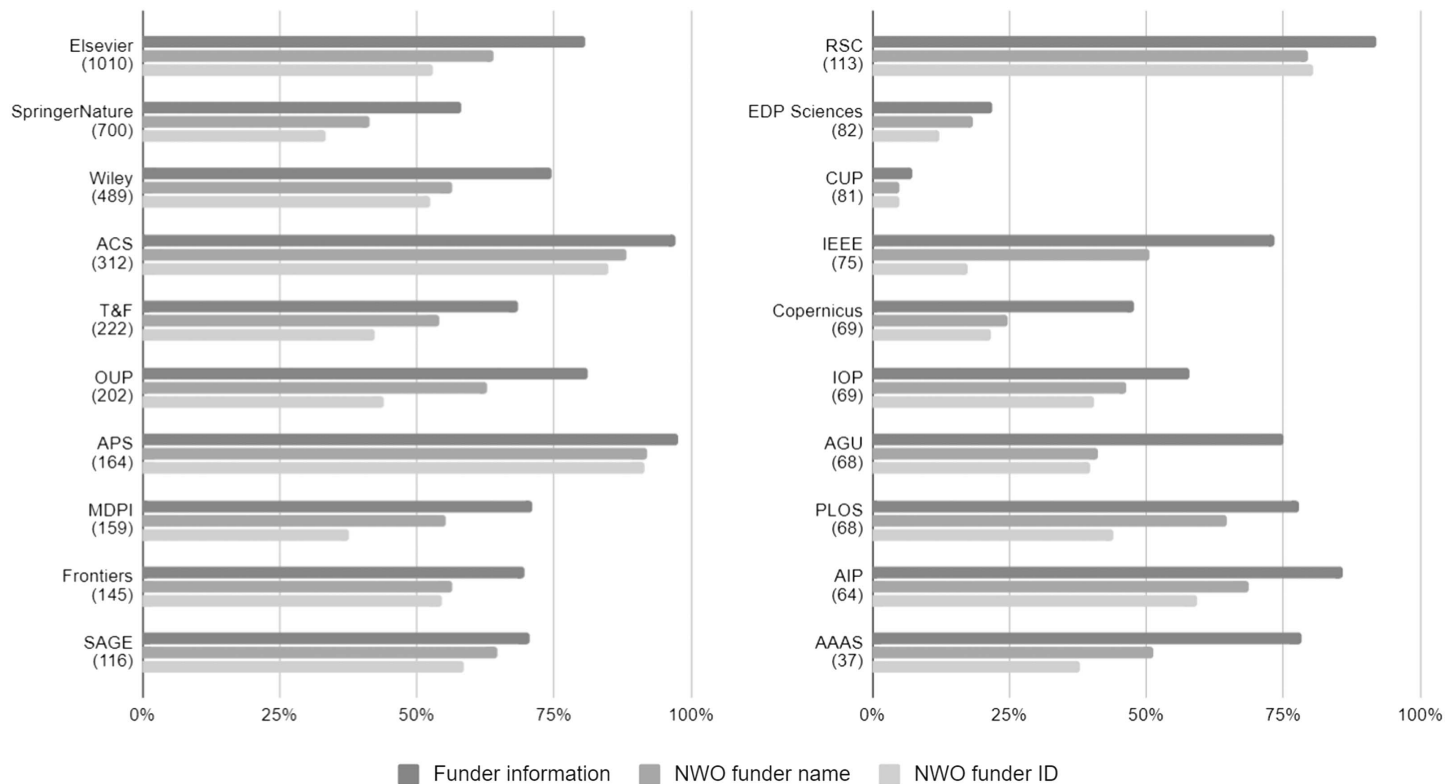

Figure 5. Variation in presence of funding information in Crossref for the 20 largest publishers in our data set.

Table 3. Availability of funding information in funding acknowledgments in papers published by PLOS and CUP

	Total number of papers in sample	Number of papers which lack funding info in Crossref ...	... of which have funding info in FA section ...	... in which NWO is acknowledged
PLOS	68	15	13	11
Cambridge University Press	81	75	62	42

correctly linking papers to external funding. A better performance of this group of full open access publishers was also expected, as one would expect that digitally native publishers have a technological advantage over large legacy publishers in collecting this kind of information.

On the other end of the spectrum we see some publishers who do not seem to provide funding information to Crossref for most of their publications, or might have only started very recently. EDP Sciences provides funding information in only 22% of cases and Cambridge University Press (CUP) for only 7% of its publications in our data set.

The fact that some publishers do not register funding information with Crossref for a significant share of their publications probably has a technical background and cannot be explained by the fact that authors do not provide that information. From a manual spot check we performed, we conclude that in most cases the information is indeed available as part of the funding acknowledgment section of the articles. For this random sample we looked at the publications by both PLOS and CUP. Table 3 shows that papers for which no funding data are available in Crossref often do contain funding information in either the acknowledgment section or footnotes of the manuscript. And in most cases NWO is also correctly identified as funder. Apparently, this information does not automatically find its way to Crossref when the papers are being registered.

3.3. Funder Data in Other Bibliographic Databases

3.3.1. Overall availability of funding metadata

Of course, Crossref is not the only source to provide funding information. WoS was the first to start collecting this information, from 2008 onwards. Scopus followed in 2013. More recently, Lens and Dimensions were launched, which both also contain funder information. Figure 6 shows how the four large bibliographic databases perform based on our data set compared to Crossref, both in coverage of DOIs (light gray bar segments) and in availability of funder information (colored bar segments).
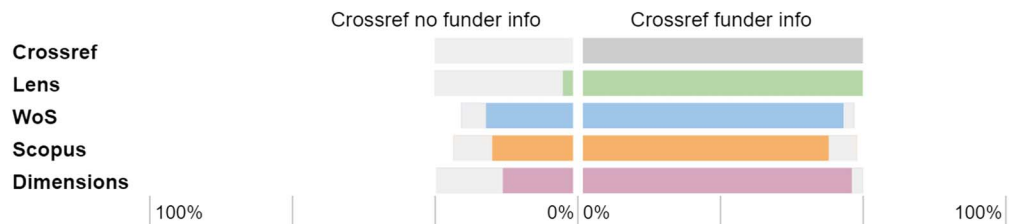

Figure 6. Performance of four bibliographic databases in providing funder information compared to Crossref. The number of publications lacking funder information in Crossref that are found in each of the other databases is indicated by the light grey bars on the left of the figure, while the number of publications having funder information in Crossref that are found in each of the other databases is indicated by the light grey bars on the right of the figure. The colored bars show the proportions of each of these sets of papers for which the other databases have funder information.

As we have seen above, 67% of DOIs in our data set contain some kind of funding information in Crossref. Lens, which takes metadata directly from Crossref, attains a similar percentage. In addition, Lens provides funding information for a small number ($n = 122$) of publications for which no funding data are available in Crossref. Funder information for these publications is derived from PubMed, which serves as an additional source of funder metadata in Lens[8].

The three commercial bibliographic databases all perform slightly less well for those records where Crossref provides funding information, but all three have funding information for a considerable number of publications that do not have funding information in Crossref.

WoS and Scopus provide funding information for, respectively, 93% and 88% of the DOIs present in the database that contain funding information in Crossref. In addition, they have funding information for over half of publications for which no funding information is available in Crossref (1,042 and 968 publications, respectively). In Scopus, this information is extracted from the acknowledgment sections of the papers, using natural language processing techniques (Baas, Schotten et al., 2020). In WoS, information from the funding acknowledgment section is enriched with information from funder repositories[9]. Both WoS and Scopus also lack a number of DOIs in our data set (419 and 292, respectively), due either to their selective coverage or to a delay in including recent publications (see below).

Dimensions is an interesting case, given the database's strong commitment to provide information about the connections between publications, awarded grants, data sets, and other outputs from the larger research life cycle. Dimensions provides funding information for nearly all records that have funding data in Crossref, which can be explained by the fact that Crossref (together with PubMed Central) figure as important "backbone" sources for Dimensions (Herzog et al., 2020). In addition, Dimensions provides information for 840 publications that lack this information in Crossref, collected from acknowledgment sections through text mining.

For all publications in our data set ($n = 5,004$), WoS still provides the most complete information regarding the source of funding (83%), followed by Dimensions (81%), Scopus (78%) and Lens (69%). In those cases where no funding information is available in Crossref, based on our data set, WoS seems to do better than Scopus and Dimensions in providing a comprehensive picture of the output funded by now.

Figure 7 shows that for all four databases under consideration in this paper, the completeness of funding data has increased over time. Interestingly, where both Scopus and Dimensions show an increase of funder information for the most recent year (2022), WoS shows a drop. This could suggest that WoS does not publish funding information immediately upon inclusion of the publications in the database but only gradually adds that to the records at a later stage.

Another observation is that both Scopus and WoS at the time of data collection (April 2022) still had a backlog in terms of overall coverage for the previous calendar year (2021) as they both miss a number of publications for which metadata is already available in Crossref ($n = 189$ and $n = 134$ respectively).

3.3.2. Differences per publisher

We also analyzed how well funding information is presented in the various bibliographic databases for publications by different publishers. This is interesting because it can tell us something about how effective publishers are in collecting this information and making it available,

---

8 https://about.www.lens.org/.
9 https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Availability-of-funding-data? (accessed June 16, 2022).
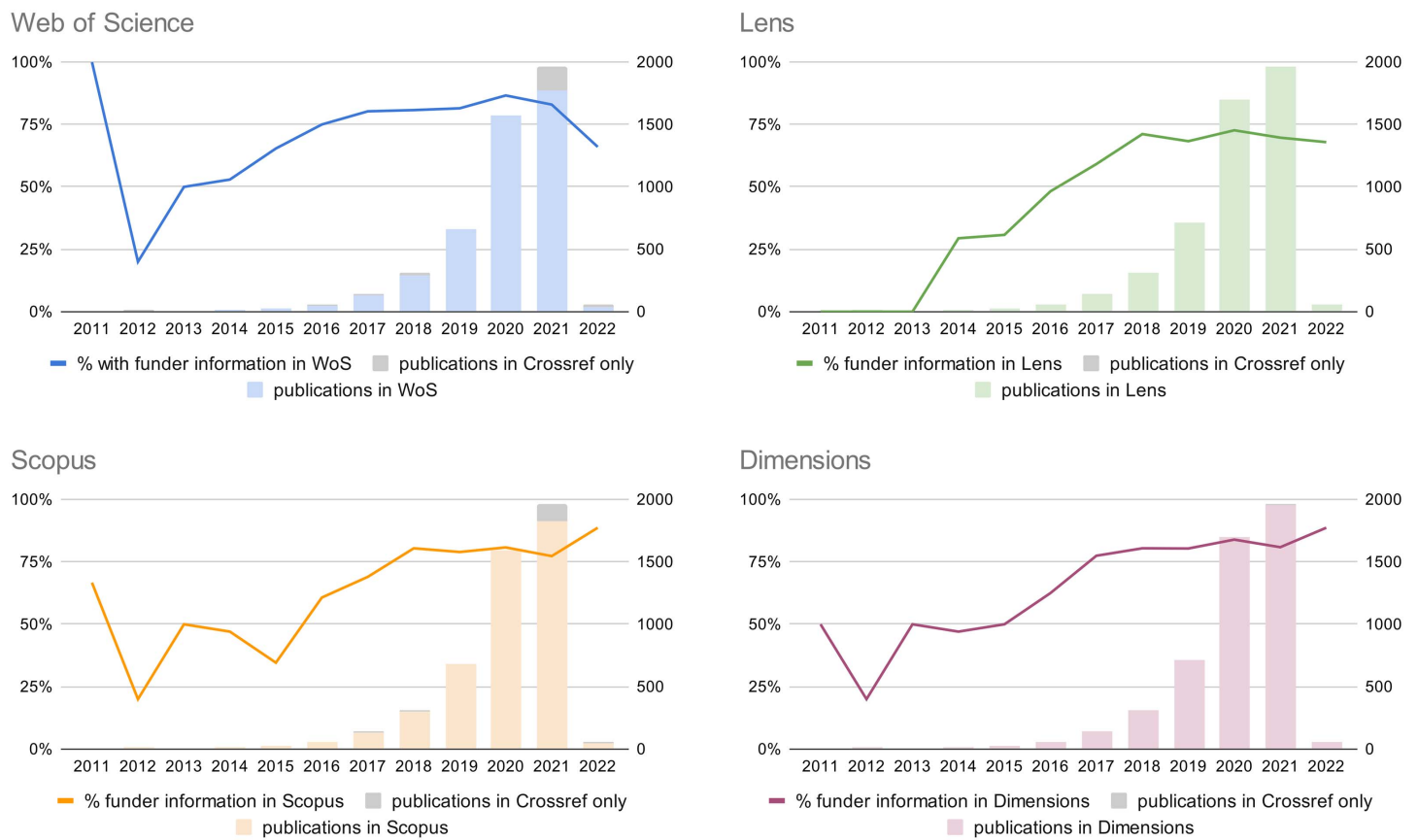
Figure 7. Crossref records (*n* = 5,004) in different databases, as well as percentage of DOIs that have funding information in that database, by publication year.

either in the full text of publications or in the metadata. Figure 8 presents the results for the 20 largest publishers in our data set, compared to the level of funding data in Crossref.

Again, we see the group of large society presses that provide funding information for nearly 100% of their publications in Crossref. These publishers perform equally well in WoS, Scopus, Lens, and Dimensions. Apparently, these publishers are able to collect this information in a very efficient way from the authors and can easily provide this information to the larger bibliographic databases.

We also see examples of publishers whose performance is mediocre when it comes to the coverage of funding data in Crossref, but for which the large bibliographic databases are able to provide funding information in sizable quantities. EDP Sciences, CUP, and, to a lesser extent, Copernicus are examples at hand. Funding information for these publishers is not well provided to Crossef. WoS, Scopus, and Dimensions, however, provide funding data for sizable numbers of publications. In the case of CUP this amounts to 58 out of 81 publications.

This, again, shows that this information is available somehow but apparently not in a format which allows the publisher to easily register these data to Crossref, or publishers choose not to deposit it to Crossref in the first place. Probably, as we have seen above with the sample of papers published by PLOS and CUP, the information is available as part of the full text of the papers and is extracted by WoS, Scopus, and Dimensions. Overall, these databases seem to employ techniques that are quite comparable in terms of performance.
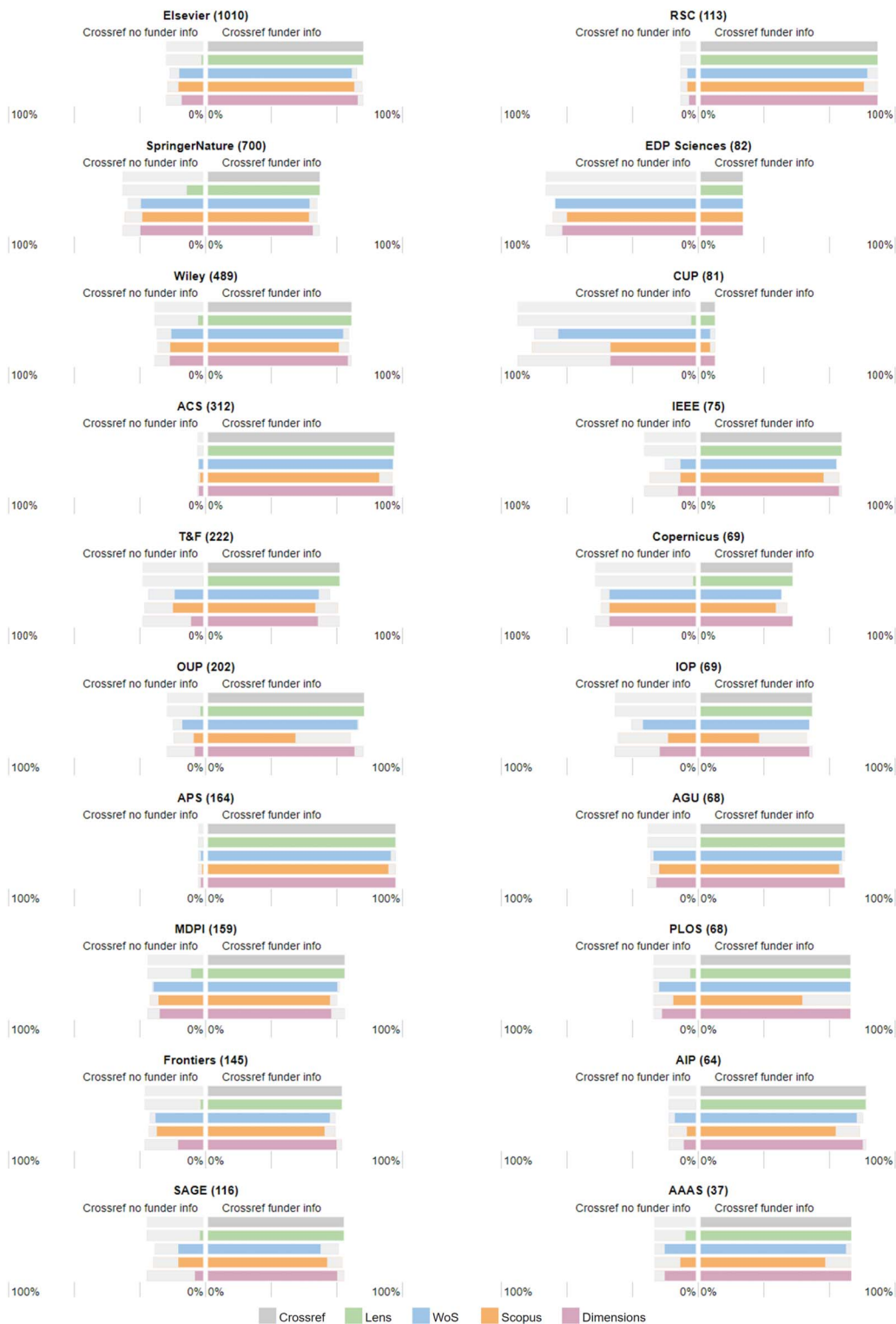
Figure 8. Performance of bibliographic databases in providing funder information compared to Crossref, broken down by publisher.

Interestingly, for publications by Taylor & Francis, SAGE, and Frontiers, Dimensions is doing less well in collecting funder information when compared to WoS and Scopus. In other cases, however, Dimensions seems to outperform its competitors. For publications with PLOS, Institute of Physics (IOP), and AAAS, for instance, Dimensions has a more comprehensive coverage than at least Scopus.

4. DISCUSSION AND CONCLUDING REMARKS

Crossref is becoming an increasingly interesting bibliometric resource, and with the start of the Fundref project in 2013 it has also become an important open source to trace research funding. Earlier research (Habermann, 2019; Hendricks, Tkaczyk et al., 2020; van Eck & Waltman, 2021) has already shown that since its start the availability of open, standardized information about the funding of publications has increased considerably. Today 25% of Crossref records contain some kind of funding information.

This study is the first to assess the completeness of that data by using a set of publications that have been reported by grantholders to be the result of external funding and therefore—in theory—funding information should be available.

We conclude that there is room for considerable improvement: 67% of publications in our data set contained funder information in Crossref. Importantly, in only 53% of cases was NWO identified by funder name and only 45% with the funder ID.

Our analysis shows that some publishers provide this information to Crossref for nearly 100% of the publications in our data set, but that there are also publishers that do not yet provide this information to Crossref. These differences, our research has shown, cannot be explained by the fact that funding information is not available for these publications. A limited manual spot check for two publishers showed that nearly all publications lacking funder information in Crossref did actually contain funding information in the acknowledgment sections of the papers.

Using data from acknowledgment sections of papers, the large commercial bibliographic databases (WoS, Scopus, and Dimensions) are able to provide funding data for a sizable proportion of publications (in our case up to 1,042 publications that did not have funding information in Crossref). Lens, by contrast, takes funder metadata directly from Crossref and other open data sources, and funder coverage closely matches that in Crossref. Interestingly, there are considerable differences in the extent to which the three commercial bibliographic databases succeed in extracting funding information for publications of different publishers. This seems to suggest that WoS, Scopus, and Dimensions differ in the extent to which they have access to the content of various publishers.

It is clear that even where funding information is available in the acknowledgments section of the paper, it is not always deposited to Crossref. Differences between publishers might be the result of the way publishers collect and process this information, including linking funder names to funder IDs. It seems quite likely that collecting funding information as part of the submission workflow is more efficient than extracting this information from funding acknowledgments sections or footnotes at a later stage. However, many publishers seem either unable to process and submit this information or choose not to do so for at least some of the articles they publish.

It has also been suggested (Mugabushaka et al., 2022) for Crossref to include the full acknowledgments section as a metadata field. While this would help make acknowledgment information (including funding information) available for analysis in this way, it would not solve issues of standardization. In addition, as the full text of acknowledgments sections could

be considered to fall under copyright, it would not be available for full reuse as are other metadata elements.

The open unrestricted availability of structured, machine-readable information about the funding of research is important for multiple reasons. To begin with, publicly funded research must be seen as a public good, the results of which ought to be open access. There is no reason why the metadata associated with publications resulting from public funding should not be openly available. From a practical perspective, publishers themselves will be interested to know how the research they publish is being funded. For funders, it will continue to be important to capture an as comprehensive as possible picture of the research they have funded, to account for the impact of their funding, to inform strategy development, and to check compliance with their open access policies. The open availability of these data may also in due course reduce the administrative burden on researchers, by preventing them from having to provide the same information multiple times in multiple places. Finally, the open availability of this information may reduce the dependency of the academic sector on third-party providers of bibliographic databases and information systems that capture the outputs of funded research, such as Researchfish.

Of course, coverage of funding metadata in Crossref will not always be sufficient, as many types of research output do not (yet) have a DOI. In addition, the question could be asked who should be the authoritative source of funding information: publishers, funders or researchers themselves? It is clear also that funders have a role to play here because making connections between research outputs and grants (not just funder names) becomes much easier if grant information from grant management systems is openly available. Therefore, a growing number of funders have started to register grant information with Crossref and attach grant IDs to their funded projects (Tkaczyk, 2022). Funders could also consider being more strict in their demands on how funding information should be structured in metadata. In fact, the funders that are part of cOAlition S already have set specific requirements with regard to the completeness and quality of metadata, including funding information[10].

Open funding data in Crossref form an important contribution to comprehensive open metadata that others can use and build upon. In recent years important steps have been taken to promote the open availability of metadata of the scholarly record. Nearly all big and medium-sized publishers provide citation data to Crossref, and increasingly have made them openly available as part of the Initiative for Open Citations.[11] Since June 2022, all citation data in Crossref are openly available by default. A rising, but still lower, number of publishers also provides access to the abstracts of the papers they publish, as promoted by the Initiative for Open Abstracts[12]. Our research shows that the open availability of funding information, while increasing, also needs improvement.

Crossref provides an increasingly interesting open data source to track the results of funded research. But for a sizable proportion of publications, information is lacking or incomplete, whereas these data seem to be available. A number of publishers therefore need to seriously step up their efforts to collect and submit these data to Crossref.

---

[10] A mandatory requirement for all publication venues is the inclusion of: "high-quality article level metadata in standard interoperable nonproprietary format, under a CC0 public domain dedication. Metadata must include complete and reliable information on funding provided by cOAlition S funders (including as a minimum the name of the funder and the grant number/identifier)." See https://www.coalition-s.org/technical-guidance_and_requirements/.

[11] https://i4oc.org/.

[12] https://i4oa.org/.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Hans de Jonge: Conceptualization, Formal analysis, Investigation, Writing—original draft, Writing—review & editing. Bianca Kramer: Conceptualization, Formal analysis, Investigation, Visualization, Writing—review & editing.

COMPETING INTERESTS

FUNDING INFORMATION

DATA AVAILABILITY

The data and code used in this study are available at https://doi.org/10.5281/zenodo.6795855 (de Jonge & Kramer, 2022):

- Data set of unique DOIs (*n* = 5,004) with collected information from Crossref and presence/absence of funder information in Lens, WoS, Scopus, and Dimensions
- List of funder name variants for NWO found in Crossref
- Google Apps Script for retrieving information from Crossref and processing Dimensions results
- R script for cleaning DOIs

REFERENCES

Álvarez-Bornstein, B., & Montesi, M. (2021). Funding acknowledgements in scientific publications: A literature review. *Research Evaluation*, *29*(4), 469–488. https://doi.org/10.1093/reseval/rvaa038

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

Costas, R., & Yegros, A. (2013). Possibilities of funding acknowledgement analysis for the bibliometric study of research funding organizations: Case study of the *Austrian Science Fund* (*FWF*). In *Proceedings of the 14th International Conference of the International Society for Scientometrics and Informetrics* (pp. 1401–1408). https://www.issi-society.org/proceedings/issi_2013/ISSI_Proceedings_Volume_II.pdf

Clements, A., Reddick, G., Viney, I., McCutcheon, V., Toon, J., … Wastl, J. (2017). Let's talk—Interoperability between university CRIS/IR and Researchfish: A case study from the UK. *Procedia Computer Science*, *106*, 220–231. https://doi.org/10.1016/j.procs.2017.03.019

de Jonge, H., & Kramer, B. (2022). Dataset: The availability and completeness of open funder metadata—Case study for publications funded by the Dutch Research Council. *Zenodo*. https://doi.org/10.5281/zenodo.6795855

Grassano, N., Rotolo, D., Hutton, J., Lang, F., & Hopkins, M. M. (2017). Funding data from publication acknowledgments: Coverage, uses, and limitations. *Journal of the Association for Information Science and Technology*, *68*(4), 999–1017. https://doi.org/10.1002/asi.23737

Habermann, T. (2019). The big picture—Has CrossRef metadata completeness improved? *Metadata Game Changers*. https://metadatagamechangers.com/blog/2019/3/25/the-big-picture-how-has-crossref-metadata-completeness-improved

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

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395. https://doi.org/10.1162/qss_a_00020

Inge, S. (2022). Researchfish apologises again as online backlash grows. *Research Professional News*. https://researchprofessionalnews.com/rr-news-uk-careers-2022-3-researchfish-apologises-again-as-online-backlash-grows/

Lammey, R. (2014). CrossRef developments and initiatives: An update on services for the scholarly publishing community from CrossRef. *Science Editing*, *1*(1), 13–18. https://doi.org/10.6087/kcse.2014.1.13

Liu, W., Tang, L., & Hu, G. (2020). Funding information in Web of Science: An updated overview. *Scientometrics*, *122*(3), 1509–1524. https://doi.org/10.1007/s11192-020-03362-3

Meddings, K. (2013). FundRef: Connecting research funding to published outcomes. *Insights*, *26*(3), 272–276. https://doi.org/10.1629/2048-7754.98

Mugabushaka, A.-M. (2020). *Linking publications to funding at project level: A curated dataset of publications reported by FP7 projects*. https://arxiv.org/abs/2011.07880

Mugabushaka, A.-M., van Eck, N. J., & Waltman, L. (2022). Funding Covid-19 research: Insights from an exploratory analysis using open data infrastructures. *arXiv*. https://doi.org/10.48550/arXiv.2202.11639

Tkaczyk, D. (2022). *Follow the money, or how to link grants to research outputs*. https://www.crossref.org/blog/follow-the-money-or-how-to-link-grants-to-research-outputs/

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