

Funding Acknowledgments in the Web of Science: Completeness and Accuracy of Collected Data

Belén Álvarez-Bornstein, Fernanda Morillo and María Bordons

Belén Álvarez-Bornstein,
IFS, Spanish National Research Council (CSIC), Albasanz 26-28, 28037 Madrid (Spain)
Phone: 34-91-602-28-94
belen.alvarez@cchs.csic.es

Fernanda Morillo
IFS, Spanish National Research Council (CSIC), Albasanz 26-28, 28037 Madrid (Spain)
Phone: 34-91-602-28-81
fernanda.morillo@cchs.csic.es

María Bordons (corresponding author)
IFS, Spanish National Research Council (CSIC), Albasanz 26-28, 28037 Madrid (Spain)
Phone: 34-91-602-28-76
maria.bordons@cchs.csic.es

Postprint version of: Álvarez-Bornstein, B.; Morillo, F.; Bordons, M. Funding Acknowledgments in the Web of Science: Completeness and Accuracy of Collected Data. *Scientometrics* (2017) 112:1793–1812. DOI 10.1007/s11192-017-2453-4

The final publication is available at Springer via
<https://link.springer.com/article/10.1007/s11192-017-2453-4>

Abstract

The study of acknowledgements as a source of data on research funding is gaining ground in science, as exemplified by the recent inclusion of this information in bibliographic databases such as Web of Science (WoS). This paper analyzes a random sample of articles published in eight thematic areas in 2014 and indexed in Wos to explore the completeness and accuracy of the database in extracting and processing funding acknowledgment data. Funding information that appears in original articles is recorded by WoS in the “funding text” field, but is also extracted to the “funding agency” and “grant number” subfields. In the extraction process, some funding information was lost in 12% of the articles and the distribution of agencies and grant numbers by subfield was not always consistent. In addition, our data suggest that funding support is often incompletely reported by authors: in about half of the articles we studied, the country of origin of the funder or the grant number(s) were not mentioned. We propose and discuss the need to develop more detailed guidelines on how to acknowledge funding. More accurate documentation of funding sources in published articles would benefit researchers, funders and journals, and enhance the reliability and usefulness of studies that focus on funding acknowledgements.

The study of acknowledgements in research publications is gaining ground in science. At present, acknowledgements are considered an important element in the scholarly communication process, forming part of the so-called “reward triangle” together with authorship and citations (Cronin & Weaver, 1995). Acknowledgements may contain expressions of gratitude to individuals, institutions or research infrastructures that have contributed to the work being reported, as well as formal recognition to research funders. Accordingly, the study of acknowledgements is of interest not only to better understand communication practices in science, but also for research policy purposes (Paul-Hus, Desrochers & Costas, 2016; Rigby, 2011).

The collection of funding acknowledgments merits special attention because it may facilitate a wide range of studies. The aims of such efforts include tracking the research output supported by funding bodies and specific grant and research programs (e.g. Boyack & Jordan, 2011; Lewison, 1994; Lyubarova, Itagaki, & Itagaki, 2009), exploring the relationship between research impact and economic support (e.g. Harter & Hooten, 1992; Lewison & Dawson, 1999; Rigby, 2013; Zhao, 2010), identifying main research funders by country (e.g. Wang, Liu, Ding, & Wang, 2012), investigating the strategic scope of a funding body (e.g. Wang and Shapira, 2011), identifying funders of the most cited papers (Giles & Councill, 2004) or studying interactions between the public and private sectors (Morillo, 2016), among other topics.

Since 2008, when funding acknowledgments were first included in Web of Science (WoS), the number of bibliometric studies focused on this source of information has increased. From studies at the micro level which analyze small units such as journals (see, for example, Cronin & Show, 1999; Salager-Meyer, Alcaraz-Ariza, Luzardo-Briseño, & Jabbour, 2011; Zhao, 2010), research has moved to studies of disciplines (see for example, Díaz-Faes & Bordons, 2014; Wang & Shapira, 2011) or countries (see, for example, Wang et al., 2012), and most recently to macro-level studies which focus on worldwide scientific output (e.g. Costas & Van Leeuwen, 2012). Regardless of the level of analysis, however, the results of these studies may be affected by technical limitations in how databases process information, as well as by shortcomings or limitations in how authors provide their acknowledgments. Concerning the latter, it has been noted that authors do not always acknowledge their funders (Costas & Yegros-Yegros, 2013; Grassano, Rotolo, Hutton, Lang, & Hopkins, 2017; Koier & Horlings, 2015), and that cultural and political issues may have an influence on the way in which authors acknowledge their support (Rigby, 2011). It has been also shown that some acknowledgments retrieved from publications may be unknown to the corresponding funders, probably because some articles are published after finalization of the project and not included in the final report submitted to the funder (Costas & Yegros-Yegros, 2013). This potential source of uncertainty supports the usefulness of publication-based studies.

The analysis of funding data is a difficult task because the data are usually included in the article’s acknowledgments, which is a non-structured section with heterogeneous content comprising technical, ethics-related or intellectual support in addition to financial assistance (Cronin, 1995). Early studies focused mainly on personal acknowledgments and usually relied on manual procedures for data analysis (e.g. Cronin, McKenzie, Rubio, & Weaver-Wozniak, 1993). More recently, the inclusion of acknowledgments in databases such as WoS (since August 2008) or Scopus (since June 2013¹) has made automatic extraction of data possible, but the heterogeneous nature of the field remains an obstacle to their systematic exploitation. Nevertheless, due to funding body requirements, funding acknowledgments are increasingly included in publications and additional efforts are needed to assess the reliability of the data collected from this source, which is crucial for the trustworthiness of subsequent analytical studies.

In the WoS database, funding acknowledgments are captured if they meet the following criteriaⁱⁱ.

- Document type: only for articles and reviews
- Indexing: included in Science Citation Index Expanded since 2008 and in the Social Sciences Citation Index since 2015
- Language: funding statement information is in English

The paragraph containing funding information in the original publication is recorded in a field named “funding text” (FX). Moreover, the funding organizations and grant numbers are extracted and recorded in two separate subfields: “funding agency” (FA) and “grant number” (GN). An illustrative example is shown in Figure 1.

Figure 1. Example of the funding acknowledgment field (FX) and subfields FA and GN in a WoS publication

FX		
We are grateful for the financial support of this work to the Spanish Ministry of Science and Innovation (project CTQ2011-28967) and to FEDER funds from the European Union		
↓		
FA-GN entries	FA	GN
FA-GN1	Spanish Ministry of Science and Innovation	CTQ2011-28967
FA-GN2	FEDER funds from the European Union	

Note. “FA-GN entry” is our own term to designate the pair of FA and GN subfields corresponding to a specific source of financial support

In principle, the extraction of funders from the full text is a valuable improvement which may enhance funding acknowledgment studies. However, a number of factors remain that constitute an obstacle to the reliability of studies for science policy purposes. The lack of standardization of funder names, which can appear in different languages, the coexistence of long, abridged and abbreviated forms, and even spelling errors, makes it difficult to obtain reliable data on the presence of funders in the acknowledgment field (Sirtes, 2013). Moreover, WoS errors in the collection and processing of FA data can be an important potential hindrance to FA analyses. Since the reliability of acknowledgment-based studies depends to a great extent on the truthfulness and accuracy of the collected data, better knowledge of how funding information is collected and processed by Thomson-Reuters (now Clarivate) is needed to assess the validity of WoS-based analyses and their usefulness for research policy studies. In this connection, a few recent studies analyzed the scope and coverage of the funding acknowledgment data collected by WoS (Paul-Hus, Desrochers, & Costas, 2016; Tang, Hu & Liu, 2017). Moreover, the recall and precision of WoS in extracting funding acknowledgments from publications has been analyzed in specific areas such as climate change (Koier & Horlings, 2015) and cancer research (Grassano et al. 2017). According to Koier and Horlings, some loss of information by WoS in the recognition of acknowledgments in the original publications was observed in 24% of the articles, while inaccuracies in the identification of individual funders appeared in another 12%. The study by Grassano et al. (2017) reported recall and precision values above 90%, although the authors suggest that differences among research domains may exist. In this context, our study seeks to improve our knowledge about different aspects of the “quality” of funding data collected by WoS through the analysis of a sample of publications from different research domains. Some preliminary data were presented elsewhere (Álvarez-Bornstein, Morillo, & Bordons, 2016).

Objectives

The main objectives addressed in this article are the following:

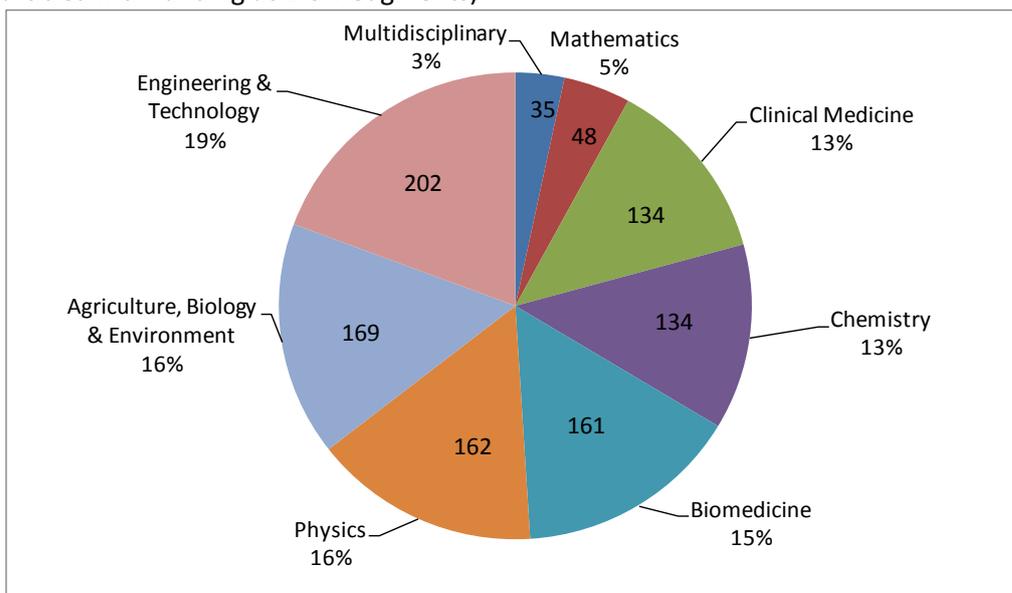
- to explore WoS completeness in extracting data from the “funding text” (FX) field as well as the accuracy of the distribution of these data by FA and GN subfield
- to analyze the extent to which funding information is complete regarding two important aspects: whether both FA and GN for a given source of support are included, and whether the country of origin of the funding is mentioned.

Our ultimate objectives are to propose recommendations intended to improve the way in which authors and databases include funding information, and in the long term, to enhance the reliability and usefulness of studies based on funding acknowledgments.

Methodology

According to WoS the scientific output of Spain-based researchers in 2014 accounted for 38,409 articlesⁱⁱⁱ published in English with funding acknowledgments (76% of the total), considering all fields except Social Sciences and Humanities. Articles not written in English and those published in Social Sciences and Humanities journals were not included because their funding acknowledgments were not captured by WoS in 2014. A stratified random sampling method was used to ensure that eight different broad thematic areas were well represented (Agriculture, Biology & Environment; Biomedicine; Clinical Medicine; Chemistry; Engineering & Technology; Mathematics; Multidisciplinary; and Physics). A random sample of articles from each area was obtained in a number proportional to the area's size compared to the population (total Spanish output), that is, a proportional allocation design was used, with a confidence level of 95% and a sampling error of 3% (Figure 2). A sample of 1045 articles was selected and used for this analysis.

Figure 2. Distribution of the sample of articles by broad thematic area (WoS, 2014) (N=1045 articles with funding acknowledgments)



The following aspects were analyzed:

1. Completeness of WoS in extracting funding data from the FX field

Manual review of the data collected by WoS in the FA and GN subfields was carried out and compared to the data included in the full FX field to obtain

- the percentage of articles in which all funders and grants mentioned by the authors in FX were included by WoS in FA and/or GN.
- the percentage of articles with any lost information, specifying whether it concerned funders, grants, or both.

2. Accuracy in the distribution of funding data in subfields FA and GN

In theory, each grant is allocated by the database to an FA-GN entry, including the name of the funder in the FA subfield and grant identification in the GN subfield. Accordingly, in “ordinary” cases (one funder = one grant), one would expect to find as many FA-GN entries as research grants in a given article, so the number of FA-GN entries might be a good proxy for the number of grants. However, we observed that some GN subfields included more than one grant, usually because they were supported by the same funder, whereas in some FA subfields more than one funder was recorded, frequently as a result of co-funded grants. Accordingly, we aimed to address the question of how the information was distributed in these “special” cases (co-funded grants and multiple grants from the same funder). For example, does WoS always include co-funded grants in a single FA-GN entry, or might they be distributed across different entries? The same question applies for multiple grants from a given funder. In order to answer this question the following measures were obtained:

- For the set of articles with co-funded grants (see the example in Figure 3), the proportion of articles with co-funders in a single FA-GN entry (e.g. article A in Figure 3), as well as the proportion of articles with co-funders in different entries (e.g. article B in Figure 3) was calculated.
- For the set of articles with more than one grant from the same funder (see the example in Figure 4), the proportion of articles in which these grants were included in a single FA-GN entry (e.g. article A in Figure 4), as well as the proportion of articles in which these grants appeared in different entries (e.g. article B in Figure 4) was calculated.

Two different goals were pursued: to determine whether the number of FA-GN entries is a good proxy for the number of grants, and to explore the consistency of WoS in their coverage of funding data in “special cases” (multiple grants from the same funder and co-funded grants).

Figure 3. Examples of the distribution of funding data by subfield in the case of co-funded grants

Article A

FX
This work was supported by the Spanish Ministerio de Economía y Competitividad/FEDER (research project CGL2011-25634/BOS)



FA-GN entries	FA	GN
FA-GN1	Spanish Ministerio de Economía y Competitividad/FEDER	CGL2011-25634/BOS

Article B

FX
This work was funded by the Spanish Ministry of Science and Innovation (SAFESPRAY project AGL2010-22304-004-04), and the ERDF (European Regional Development Fund)



FA-GN entries	FA	GN
FA-GN1	Spanish Ministry of Science and Innovation (SAFESPRAY project)	AGL2010-22304-004-04
FA-GN2	ERDF (European Regional Development Fund)	

Figure 4. Examples of the distribution of funding data by subfield in the case of multiple grants from the same funder

Article A

FX
We are grateful for the financial support of this work to the Spanish Ministry of Science and Innovation through projects CTQ2011-28967 and CTQ2012-30101.



FA-GN entries	FA	GN
FA-GN1	Spanish Ministry of Economy and Competitiveness	AGL2007-65661, AGL2009-12485-C03-01

Article B

FX
This research was supported by US Department of Agriculture Current Research Information System Project 3655-21000-048-00D and a US Department of Agriculture Specialty Crop Research Initiative grant (project number 2011-51181-30661) to Y.W.



FA-GN entries	FA	GN
FA-GN1	US Department of Agriculture Current Research Information System Project	3655-21000-048-00D
FA-GN2	US Department of Agriculture Specialty Crop Research Initiative grant	2011-51181-30661

3. Completeness of the data in FA-GN entries

For the purposes of this study, we consider “completely identified funding grants” as those which were identified by both a funder name and a grant number. If only one of these elements was provided, we considered them as “incompletely identified funding grants”. To investigate this issue, the following measures were obtained:

- the percentage of FA-GN entries which showed only FA data (GN was empty)

- the percentage FA-GN entries which showed only GN data (FA was empty)
- the percentage of FA-GN entries with both FA and GN data

4. Identification of the funders' country of origin

Although we considered a grant to be “completely identified” if data were provided for both the FA and GN subfields, the country of origin of the funder(s) can be extremely important to avoid confusion between funders with the same name in different countries (e.g. there is a National Science Foundation in USA, but also in other countries such as Switzerland). The fact that authors very often tend to translate the original name of their funding bodies into English in articles written and published in English increases the likelihood of confusion. Moreover, the chance of misunderstanding increases if only an acronym is mentioned. Accordingly, we calculated the percentage of FA in which the country of origin of the funding body was clearly stated as a country name or an adjectival form. Thus examples of clear identification include “Swiss National Science Foundation” or “National Science Foundation, Switzerland”. This indicator was calculated separately for Spanish and foreign funders, to explore potential differences in the tendency of authors to mention the funder's country of origin. As a preliminary step to obtain these data, the FA subfield was examined manually and funders were characterized as “foreign” or “Spanish” as appropriate. Doubtful cases were resolved through web searches.

The statistical analyses used to study potential between-area differences in the behavior of funding-related variables included contingency tables, and Kruskal-Wallis and Mann-Whitney tests for non-parametric variables ($P < 0.05$) (SPSS version 22).

Results

For the 1045 articles we analyzed, a total of 2924 FA-GN entries were collected, i.e. a mean of 2.8 FA-GN entries per article, ranging from 2.5 in Mathematics to 3.4 in the Multidisciplinary area (see detailed data in Appendix 1). Differences between areas were statistically significant ($P < 0.05$).

Completeness of WoS in Extracting Funding Data from the FX Field

The rate of completeness of WoS in extracting funding information was quite high, since all funders and grants included in the FX field were extracted to the FA and/or GN subfields in 88% of the articles. The rate of completeness ranged from 94% in the Multidisciplinary area to 86% in Engineering/Technology and Biomedicine, but between-area differences were not statistically significant. We wondered whether longer FX fields, which supposedly contain more complex texts, were less well covered than shorter ones. This was indeed the case, since the subset of articles with lost data had significantly longer FX fields (higher number of characters) than articles for which all the information was collected ($Z = -4.023$, $P < 0.001$).

Concerning lost information, a funding grant was entirely lost (neither the funder nor the grant number provided in the full text was shown in FA/GN) in 75 articles (7.2% of articles), and was partially lost in other cases in which only the funder (31 articles, 3%) or only the grant number (38 articles, 3.6%) was missing. Overall, information was missing in 13.2% of the articles (138 articles)^{iv}.

Accuracy in the Distribution of Funding Data by Subfield

Co-funded grants were found in 214 articles (20.5%). In 24.8% of them, different co-funders were recorded in a single FA-GN entry while in 75.2% they were recorded in different entries (Table 1).

Multiple grants from the same funder were detected in 341 articles (32.6%) (Table 1). These grants tend to be mentioned together in a single FA-GN entry (75.1%) (see, for example, Article A in Figure 4) but sometimes they were separated in different FA-GN entries (38.4%) (see Article B in Figure 4). In a small proportion of articles (13.5%) multiple grants mentioned in a single entry co-existed with other(s) separated in different entries.

Table 1
Distribution of funding data in two special cases

Special cases	No. Articles (%)
Articles with co-funded grants	
Two or more funders in a single FA-GN entry	53 (24.8%)
Funders in different FA-GN entries	161 (75.2%)
Total	214 (100%)
Articles with multiple grants from a given funder	
Two or more grants in a single FA-GN entry	256 (75.1%)
Grants in different FA-GN entries	131 (38.4%)
Total ^a	341 (100%)

^aThe sum is greater than the total due to overlapping papers.

Focusing on thematic areas, the proportion of articles with co-funded grants ranged from 14.3% in the Multidisciplinary area to 27.6% in Chemistry (with no significant differences by area), whereas the rate of multiple grants from a given funder ranged from 14.6% in Mathematics to 41.6% in Engineering/Technology ($P < 0.05$). Interestingly, no between-area differences were observed in the way these data were distributed across subfields by WoS in the case of co-funded grants (Figure 5) or in the case of multiple grants from the same funder (Figure 6).

Figure 5. Distribution of funding data by thematic area in articles with co-funded grants

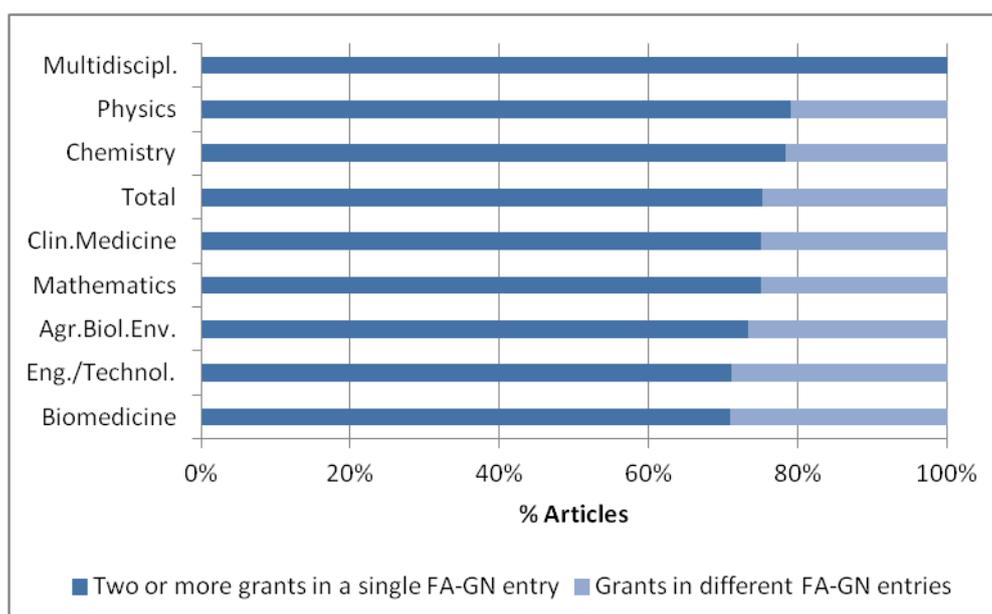
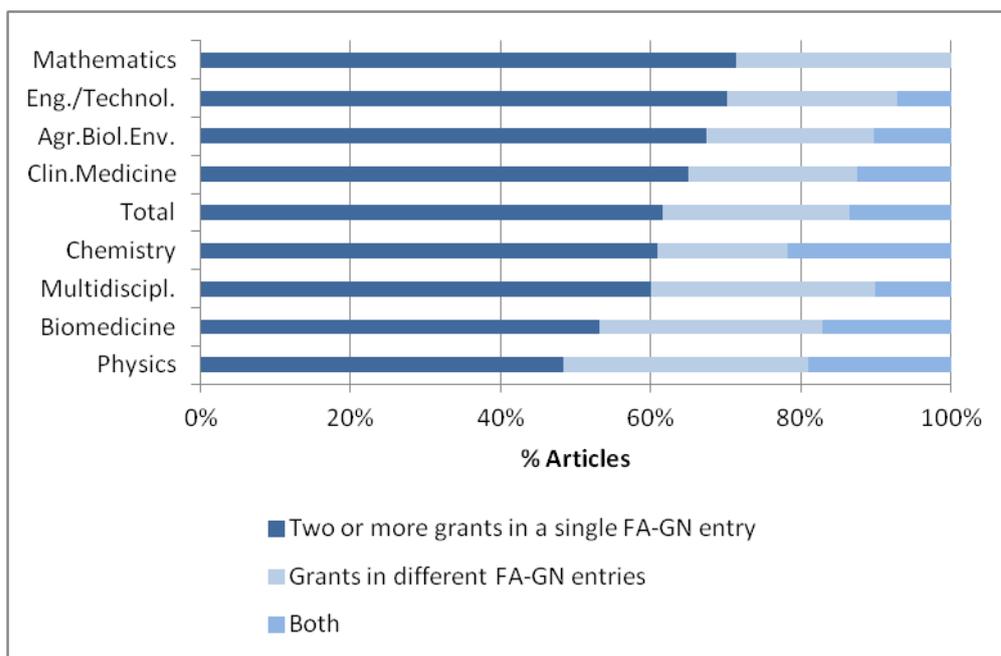


Figure 6. Distribution of funding data by thematic area in articles with multiple grants from the same funder



Finally, different types of misallocation of data were observed, such as including program or project names instead of the funder in the FA subfield (9.2% of the FA subfields and 20.1% of the articles contained at least one program or project instead of a funder), or splitting a funder’s name across two or more FA-GN entries as if there were more than one funding agency (observed in 2.7% of the articles) (see Figure 7 for an example).

Figure 7. Example of a single funder incorrectly split into two FA-GN entries

FX		
This work is financially supported by the Secretary General for the Sea of the Spanish Ministry of Agricultural, Land and Marine Resources (MARM), by order ARM/1193/2009		
↓		
FA-GN entries	FA	GN
FA-GN1	Spanish Ministry of Agricultural	
FA-GN2	Land and Marine Resources (MARM)	ARM/1193/2009

Completeness of Data Collected in FA-GN Entries

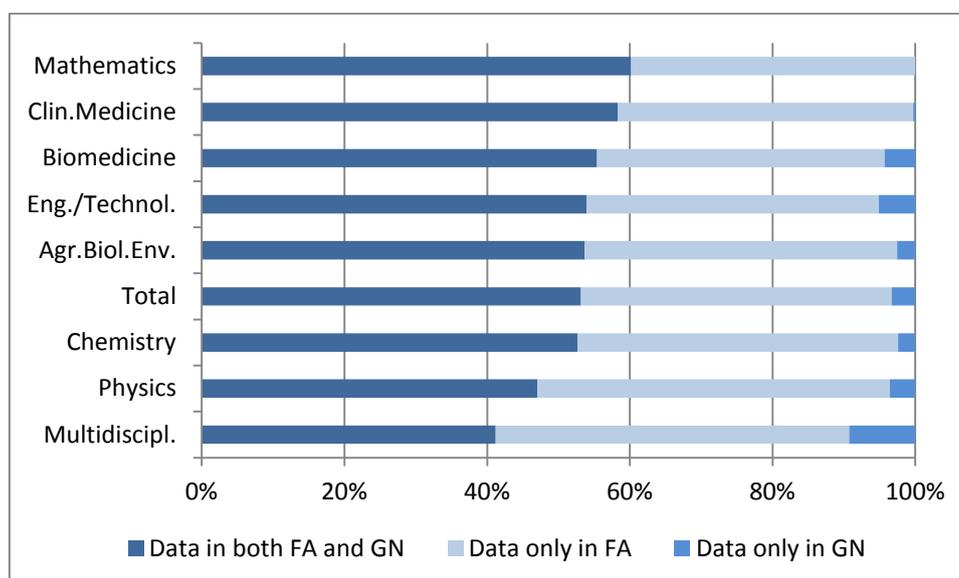
Around half of the FA-GN entries (53.1%) were “completely identified”, that is, they included at least one funder and one grant number in the corresponding subfields. However, funders were present in almost all the FA-GN entries (96.6%), whereas grant numbers appeared in only 56.5% of the cases. In 43.5% of the entries only funders were provided, whereas grants appeared as the only information in only 3.3% of the entries (Table 2).

Table 2
Completeness of FA-GN entries

	No. FA-GN entries (%)
Data in both subfields (FA and GN)	1554 (53.1%)
Data only in FA	1273 (43.5%)
Data only in GN	97 (3.3%)
Total	2924 (100%)

Turning to thematic areas, the highest proportion of complete FA-GN entries (with data in both FA and GN subfields) was observed in Mathematics (60.2%), whereas the lowest proportion was found in the Multidisciplinary area (41.2%). Interestingly, this latter area also had the highest share of entries that contained only grant data (9.2%) (Figure 8). Overall, between-area differences in the completeness profile of FA-GN entries were statistically significant ($P < 0.001$).

Figure 8. Completeness of FA-GN entries by thematic area



Identification of the Funder's Country of Origin

Funders included in the FA subfield were classified as Spanish (61.3%) or foreign (36.9%) according to their country of origin (3.6% of funders were not classified because there were no data in FA or the data included did not correspond to any funder). For this classification, the funders included in FA were examined to look for the following evidence: first, a country name (e.g. "Spain" or "Italy") or an adjectival form (e.g. "Spanish" or "Italian") in the funder's name; second, cities or regions which could be ascribed unequivocally to countries; and third, well-known funders or programs in Spain or other countries which could be ascribed unequivocally to their corresponding country or supranational region.

As seen in Table 3, the names of around 42% of Spanish funders and 50% of foreign funders included a country name or an adjectival form. In most of the remaining cases, we could determine the country of origin of the funder with variable degrees of difficulty and precision. In many cases the country was identifiable unequivocally when cities were mentioned,

although caution was needed because of homonymous cities in different countries (for example, Cordoba is a city in Spain and also in Argentina). With regard to funding from Spain, the proportion of funders' names that included a region but not a country was high (22.5%) probably because of research funding bodies administered by regional governments (e.g. the Generalitat Valenciana in the autonomous region of Valencia in Spain). Finally, some well-known funding institutions and programs could be ascribed to countries or supranational regions unequivocally. Examples included funders such as ANR in France, MIUR in Italy or MICINN in Spain^v, as well as programs such as the European Framework Program in the European Union. We identified approximately 9.5% of Spanish funders and 22.2% of foreign funders by means of web searches.

Table 3
Identification of the national or foreign origin of funding

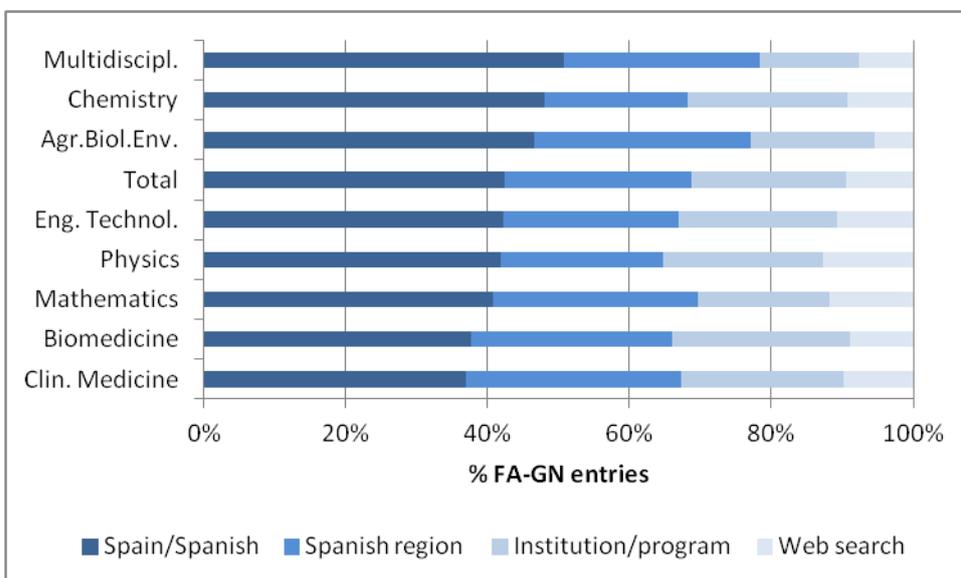
Type of funding	No. Funders (% by total)	No. Funders (% by type of funding)
1. Funding from Spain	1792 (61.3%)	
"Spain" or "Spanish" in FA		761 (42.5%)
Spanish region in FA (NUTS 2)		404 (22.5%)
Spanish cities in FA		69 (3.8%)
Institutions/programs in FA		388 (21.6%)
Other (identified by web search)		170 (9.5%)
		(100%)
2. Foreign funding	1080 (36.9%)	
Country/region included in FA ^a		542 (50.1%)
Institutions/programs in FA		299 (27.7%)
Other (identified by web search)		239 (22.2%)
		(100%)
3. Non-identified type of funding	110 (3.6%)	
Total ^b	2924 (100%)	

^a Funding from supranational entities such as the European Union was included here if the foreign source of the funding could be determined from the recorded data. ^b The sum is greater than the total number of entries due to co-funding by both Spanish and foreign funders in some entries.

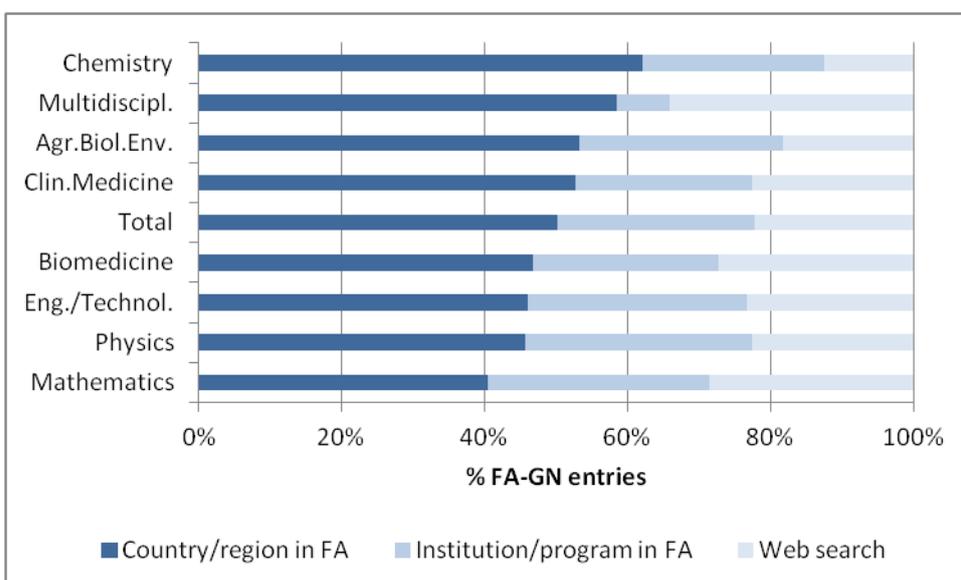
We observed some differences among thematic areas in how the national or foreign origin of funders was identified (Figure 9). The behavior of the Multidisciplinary area deserves special mention: on the one hand, a country name appeared in a high share of entries; on the other hand, the national or foreign origin of the funder was identifiable from the name of the institution or program less often than in the remaining areas (Figures 9a and 9b). In fact, in this area web searches were needed to identify the type of funder in many cases, especially for foreign funders. Overall, between-area differences were more significant for foreign funders, where the country or region was identifiable for around 60% of funders in Chemistry but only 40% of those in Mathematics.

Figure 9. Identification of the national or foreign origin of funding by thematic area

a) Funding from Spain



b) Foreign funding



Discussion and Conclusions

This study focuses on the funding acknowledgment section of a sample of articles covered by the WoS to analyze the completeness of the database in extracting data from the FX field to the FA and GN subfields, its accuracy in the distribution of data by subfields, and the completeness of the collected data. Our analysis shows that there is room for improvement in both the completeness in the data extracting process (some funding information was lost in

12% of the articles), and in the accuracy of the subfield distribution of the data, which was not always consistent.

Completeness of WoS in Extracting Funding Data

Our findings on the extraction of funding data from the FX subfield are consistent with those described by Koier and Horlings (2015), who observed some loss of information in the extraction of funders from the acknowledgment section in 12% of the articles in a study of the scientific output of two Dutch programs in climate change research. Interestingly, these authors also showed some loss of information in the previous stage, i.e. in the ability of WoS to recognize and retrieve funding acknowledgments from original papers. The fact that the location of funding acknowledgments in papers may vary from a general acknowledgment section to a specific funding section or even within a footnote hinders their recognition and subsequent extraction (Grassano et al. 2017). Further analysis of this issue was beyond the objectives of our study, but should be kept in mind in future studies of the contents of the acknowledgments.

Accuracy in the Distribution of Funding Data by Subfield

Our findings pinpoint some issues with inconsistency in the distribution of funding data by subfield. On one hand, the names of projects and programs are sometimes the only information included in the FA subfield, where it is assumed that funders will appear. In some cases this occurs because the funder's name is not provided by the authors, as has also been detected in other cases in which the funder was mentioned, but was included by WoS in a separate entry. On the other hand, we found inconsistencies in the way that co-funded grants and multiple grants from the same funder were recorded, since sometimes they appeared in a single FA-GN entry, whereas on other occasions they were disaggregated into different entries. This means that the basic structure of "one FA-GN entry = one funding grant" was not always maintained. Therefore the number of FA-GN entries is not an accurate measure of the number of sources of funding, which is an important consideration from a bibliometric perspective. In fact, our data suggests that the number of FA-GN entries is a better proxy for the number of funders than for the number of grants, since the latter may be underestimated due the presence of multiple grants within a given FA-GN entry (in 256 articles comprising 24% of the 1045 papers analyzed here). Although the presence of multiple funders in a given FA has been also detected, and may reduce the validity of the number of FA-GN entries as a measure of the number of funders, this type of co-occurrence was less common (in 53 articles, i.e. 5% of the 1045 papers we analyzed).

Completeness of Collected Data

With regard to the rate of completeness of the FA and GN subfields, a given funder was almost always included in the FA subfield (97% of FA-GN entries) whereas specific grant information appeared in only about half of the entries. This suggests that authors very often provide incomplete information about their funding grants, giving priority to the funder over the grant number. However, other factors may also contribute to the incompleteness in the FA and GN subfields:

- Some data may be lost by the database during the extraction process from the original paper or from the FX field. The first issue was not analyzed in this study, although we did find that the grant number was lost in the extraction from the FX field in around 11% of the papers (in 3.6% of the papers only the GN information was lost; in 7.2% both GN and FA data were lost).

- Financial support might not be associated to a specific grant number in some cases, such in those involving “core funding” from institutional resources. Nonetheless, this type of funding is rarely acknowledged by authors (Grassano et al., 2017).
- In some cases such as in co-funded grants, the basic structure “one FA-GN = one funding grant” was not preserved. Co-funders may be distributed in different FA-GN entries, whereas the grant number appears only once, so the additional GN subfield/s remain empty (in around 14% of the papers in our study).

These factors per se, however, cannot explain the high percentage of empty GN subfields, which in many cases can be attributed to the omission of information by the author. Our findings thus suggest a need to encourage authors to include more complete data in their funding acknowledgments.

National and Foreign Research Funders

The classification of research funders under the category of “national” or “foreign” was a laborious task because very often the country of the funder was not mentioned in the acknowledgments: it was missing for 58% of Spanish funders and 50% of foreign funders. If the country was not specified, the geographic origin of funding was traced and identified in many cases on the basis of geographic data (regions, cities) and well-known national and international institutions or programs. Web searches were crucial for identifying the funder’s country of origin in the remaining 10-20% of the FA entries. In the light of these results, we stress that the lack of information about the funder’s country can be misleading, since a given funder may have the same or a very similar name in different countries. In this connection, incomplete or inaccurate information can hinder automatic counts and increase the probability of erroneous allocations. We suggest that the inclusion by authors of the country of origin of each research funder is highly desirable.

Differences by Thematic Area

Our data reveal that the completeness of WoS in covering funding information did not differ significantly among areas, whereas the completeness of FA-GN entries did vary, with the share of complete entries ranging from 41% in the Multidisciplinary area to 60% in Mathematics. This variation may be related to a number of factors such as differences in the number and/or diversity of funders by area, which might influence how authors include funding data as well as how the data are then collected by WoS. In fact, Mathematics was the area with the lowest average number of funders per paper, a situation which probably makes it easier for authors in this area to compose the funding acknowledgments, which are subsequently more likely to be collected accurately by the database. The Multidisciplinary area stands at the other end of the spectrum, with the greatest average number of funders per paper and the lowest share of complete FA-GN entries. In addition, between-area differences in the frequency of co-funded grants and multiple grants from a given agency might also influence the rate of completeness of FA-GN entries, since we have observed some degree of incompleteness in these situations. In summary, the main patterns we found for our entire sample of papers emerged at the level of specific thematic areas, although in some cases small differences between areas were detected which can probably be attributed to the varying degrees of complexity of funding environments in different areas.

Future Actions and Measures

Both the completeness and the accuracy of the process of funding data extraction are curtailed, we believe, by two important factors: a) the fact that in many journals funding acknowledgments are included in a general “Acknowledgments” section, mixed with other

types of data such as personal and institutional acknowledgments or conflicts of interest; and b) the incomplete and/or unclear way in which many authors provide funding acknowledgments in their articles. Concerning the first issue, it is clear that using separate headings to collect and identify funding data, conflicts of interest and other types of acknowledgment in journal articles is a helpful measure. In fact, some journals have adopted this practice for the sake of clarity and to avoid misunderstandings (see for example, *Virology Journal*, <https://virologyj.biomedcentral.com/>). The use of a specific heading for sources of funding (called “Funding”, “Financial support”, “Financial data”, etc.) has already been suggested as a way to facilitate the correct identification of funding data (Henderson, Howard, & Wilkinson, 2003).

Regarding the incompleteness of data included by authors, we observed different examples in this study, such as the non-inclusion of grant numbers or the inclusion of the program name instead of the name of the corresponding funder. This suggests that authors need clearer guidelines concerning both the information that should be included and the format in which it should be provided. At present, major funders require grant-holders to acknowledge the support received, but instructions about how to present this information are frequently vague. That is the case, for example, at the Ministry of Economy and Competitiveness in Spain, which funds the Spanish National Plan for Research (see, for example, the main call for research project proposals, BOE-A-2013-10258, which mentions only the need to acknowledge support from the Ministry). More specific statements can be observed in other countries such as Canada (see, for example, <http://canada.pch.gc.ca/eng/1428491994616>) and the USA (<https://grants.nih.gov/grants/acknow.htm>) or supranational organizations such as the European Union (COST, 2015). Unfortunately, funders often mention only the requirement to acknowledge funding, but provide no specific advice or instructions on the exact content or wording.

Our data suggest that authors need clearer guidelines on aspects such as the type of information to include (funder, grant number), the format and notation of the funder’s name (full name, acronym, country), and the desirable format and notation of the information in cases that involve research support from multiple funders or through multiple grants. Authors should be encouraged to use a standardized form of the funders’ name in their acknowledgments, since the lack of standardization has been recognized as an important problem in previous studies (Sirtes, 2013). Consensus between research funders and publishers would be desirable to ensure that both actors request the same information from authors. In this connection, an interesting initiative was launched by the Research Information Network in the UK (Research Information Network, 2008), as the result of discussions with UK funders and publishers, to offer recommendations for authors and publishers regarding the content of funding acknowledgments. This group proposes that research articles should have a funding acknowledgment in the form of a sentence as follows:

“This work was supported by the Medical Research Council [grant number xxxx].”

In other words, they suggest including the funder’s name written out in full, followed by the grant number in square brackets. In the case of multiple funders and grants, RIN proposes the following format:

“This work was supported by the Wellcome Trust [grant numbers xxxx, yyyy]; the Natural Environment Research Council [grant number zzzz]; and the Economic and Social Research Council [grant number aaaa].”

This is an interesting example of notation, which could also be improved with the inclusion of the funder's country of origin (if applicable). This information is relevant given the increasing internationalization of research and the existence of funder homonyms in different countries. In fact, having detailed guidelines for funding acknowledgments at the national or even international level can be an important step forward in enhancing the collection of funding data.

From the perspective of bibliographic databases, standardization of funding data is essential to enable information retrieval. Useful strategies to achieve this goal include the creation of lists with standardized names of funders and/or including specific notation to designate grants. Among the latter, the *Pubmed* system is noteworthy in that it includes grants and contracts according to the notation: grant number/funder/country. Concerning standardized lists of funders, one collaborative project worth highlighting is *FundRef* (<http://ftp.crossref.org/fundref/>), launched by CrossRef in late May 2013 in collaboration with Elsevier and other partner organizations. Interestingly, it seeks to obtain a standardized list of funders worldwide, which is incorporated into journals' manuscript tracking system so that authors can select their funder from the list and provide additional information if needed (e.g. grant numbers). This tool would ensure the use of standardized names for all funders included in the system, and would be updated periodically.

We suggest that further efforts are needed to increase the accurate disclosure of funding sources by authors and the capacity of databases to collect this information. Initiatives designed to achieve these aims will benefit researchers (who need more assistance at the time of writing their acknowledgments), funders (to better track the outcomes of the research they fund) and publishers (which will gain from uniform information across articles and journals).

Acknowledgments

This research is supported by the Ministerio de Economía y Competitividad (MINECO), Spain (research project CSO2014-57826-P and predoctoral contract BES-2015- 073537). We are grateful to Adrián Arias for his contribution in the design of the sample of articles analyzed here, to the ACUTE team for its support in processing data and to K. Shashok for improving the use of English in the manuscript.

References

- Álvarez-Bornstein, B., Morillo, F., & Bordons, M. (2016). Accuracy and completeness of funding data in the Web of Science. In 21st International Conference on Science and Technology Indicators. Valencia, Spain: European Network of Indicator Designers (ENID) and INGENIO.
- Boyack, K.W., & Jordan, P. (2011). Metrics associated with NIH funding: a high-level view. *Journal of the American Medical Informatics Association*, 11, 18, 423e431. doi: 10.1136/amiajnl-2011-000213 423
- COST Association (2015). Guidelines for the dissemination of COST Action results and outcomes. Belgium.
- Costas, R., & Leeuwen, T. N. (2012). Approaching the “reward triangle”: General analysis of the presence of funding acknowledgments and “peer interactive communication” in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8), 1647-1661. doi: 10.1002/asi.22692
- Costas R. & Yegros-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 J.Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, &

- Moed, H. (Eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference* (pp. 1401-1408). Vienna, Austria: Austrian Institute of Technology.
- Cronin, B., McKenzie, G., Rubio, L., & Weaver-Wozniak, S. (1993). Accounting for influence: Acknowledgments in contemporary sociology. *Journal of the American Society for Information Science*, 44(7), 406-412.
- Cronin, B., & Weaver, S. (1995). The praxis of acknowledgment: from bibliometrics to influmetrics. *Revista Española de Documentación Científica*, 18(2): 172-177.
- Cronin, B. (1995). *The Scholar's courtesy: the role of acknowledgments in the primary communication process*. Los Angeles: Taylor Graham.
- Cronin, B., & Shaw, D. (1999). Citation, funding acknowledgement and author nationality relationships in four information science journals. *Journal of Documentation*, 55, 4, 402-408.
- Díaz-Faes, A. A., & Bordons, M. (2014). Acknowledgments in scientific publications: Presence in Spanish science and text patterns across disciplines. *Journal of the Association for Information Science and Technology*, 65(9), 1834-1849. doi: 10.1002/asi.23081
- Giles, C.L., & Councill, I.G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 51, 17599–17604.
- Grassano, N., Rotolo, D., Hutton, J., Lang, F., & Hopkins, M. (2017). Funding data from publication acknowledgements: coverage, uses and limitations. *Journal of the Association for Information Science and Technology* (forthcoming). doi: 10.1002/asi.23737
- Harter, S.P., & Hooten, P.A. (1992). Information science and scientists: JASIS, 1972-1990. *Journal of the American Society for Information Science*, 43, 9, 583-593.
- Henderson, C., Howard, L., & Wilkinson, G. (2003). Acknowledgement of psychiatric research funding. *British Journal of Psychiatry*, 183, 273-275. doi: 10.1192/bjp.183.4.273
- Koier, E., & Horlings, E. (2015). How accurately does output reflect the nature and design of transdisciplinary research programmes? *Research Evaluation*, 24, 37-50. doi: 10.1093/reseval/rvu027
- Lewison, G. (1994). Publications from the European Community biotechnology Action Program (BAP): multinationality, acknowledgment of support and citations. *Scientometrics*, 31, 125-142. doi: 10.1007/BF02018556
- Lewison, G., & Dawson, G. (1998). The effect of funding on the outputs of biomedical research. *Scientometrics*, 41, 1-2, 17-27. doi: 10.1007/BF02457963
- Lyubarova, R., Itagaki B.K., & Itagaki, M.W. (2009). The impact of National Institutes of Health funding on U.S. cardiovascular disease research. *PLoS ONE*, 4(7), e6425. doi: 10.1371/journal.pone.0006425
- Morillo, F. (2016). Public-private interactions reflected through the funding acknowledgements. *Scientometrics*, 108(3), 1193-1204. doi: 10.1007/s11192-016-2032-0
- Paul-Hus, A., Desrochers, N., & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgment data in Web of Science. *Scientometrics*, 108, 167-182. doi: 10.1007/s11192-016-1953-y
- Research Information Network (2008). *Acknowledgement of Funders in Scholarly Journal Articles Guidance for UK Research Funders, Authors and Publishers*. Retrieved from www.rin.ac.uk/system/files/attachments/Acknowledgement-funders-guidance.pdf
- Rigby, J. (2011). Systematic grant and funding body acknowledgement data for publications: new dimensions and new controversies for research policy and evaluation. *Research Evaluation*, 20(5), 365-375. doi: 10.3152/095820211X13164389670392

- Rigby, J. (2013). Looking for the impact of peer review: does count of funding acknowledgements really predict research impact? *Scientometrics*, 94, 57–73. doi: 10.1007/s11192-012-0779-5
- Salager-Meyer, F.; Alcaraz-Ariza, M.A.; Luzardo-Briceño, M.; Jabbour, G. (2011). Scholarly gratitude in five geographical contexts: a diachronic and cross-generic approach of the acknowledgment paratext in medical discourse (1950-2010). *Scientometrics*, 86, 763-784. doi: 10.1007/s11192-010-0329-y
- Sirtes, D. (2013). Funding acknowledgements for the German Research Foundation (DFG). The dirty data of the Web of Science database and how to clean it up. In J.Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & Moed, H. (Eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference* (pp. 784-795). Vienna, Austria: Austrian Institute of Technology.
- Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: queries and caveats. *Journal of the Association for Information Science and Technology* (forthcoming). doi: 10.1002/asi.23713
- Wang, J., & Shapira, P. (2011). Funding acknowledgement analysis: an enhanced tool to investigate research sponsorship impacts: the case of nanotechnology. *Scientometrics*, 87, 563-586. doi: 10.1007/s11192-011-0362-5
- Wang, X., Liu, D., Ding, K., & Wang, X. (2012). Science funding and research output: a study on 10 countries. *Scientometrics*, 91, 591-599. doi: 10.1007/s11192-011-0576-6
- Zhao, D. (2010). Characteristics and impact of grant-funded research: a case study of the library and information science field. *Scientometrics*, 84, 293-306. doi: 10.1007/s11192-010-0191-y

ⁱ Funding information is captured if it is mentioned in the acknowledgment section of the article and if the funding body is included in the FundRef ontology: <http://www.crossref.org/fundref/>.

ⁱⁱ Information provided by Thomson Reuters Technical Support Unit (October 2015).

ⁱⁱⁱ Articles and reviews, hereafter referred to as “articles”.

^{iv} There is some overlap between categories because grants that were entirely and partially lost may coexist in a given article.

^v ANR stands for L'Agence nationale de la recherche in France; MIUR stands for the Ministero dell'Istruzione, dell'Università e della Ricerca in Italy; MICINN stands for the Ministerio de Ciencia e Innovación in Spain.