

Public funding accountability: a linked open data-based methodology for analysing the scientific productivity and influence of funded projects

Antonio Perianes-Rodríguez¹ · Carlos Olmeda-Gómez¹ · Natalia R. Delbianco² · Maria Cláudia Cabrini Grácio²

Received: 13 August 2023 / Accepted: 15 February 2024 © The Author(s) 2024

Abstract

Although funding acknowledgements (FAs) have been around for nearly three decades, there are not yet enough theoretical and practical studies of them to enable FAs to be considered a consolidated area of research. Fortunately, newly published findings and promising data sources presented in recent years have helped better our understanding of the process of scientific creation and communication and provide evidence of the importance of FAs. This paper seeks to help demonstrate the crucial role FAs play in evaluating research funding's performance. A methodology based on the use of linked open metadata from diverse sources is presented for this purpose. The methodology highlights the important work analysts do to increase the accuracy, solidity, and diversity of the results of FA-based quantitative studies by gathering and analysing the data furnished by funding organisations. Lastly, the projects funded by the Spanish National Science and Research Agency from 2008 to 2020 are evaluated to verify the method's usefulness, robustness, and reproducibility. Also, a new unit of analysis is introduced, funders, to create a new type of cooccurrence network: co-funding. In conclusion, funding agencies' experts and analysts will find that this methodology gives them a valuable instrument for boosting the quality and efficacy of their activities, complying with transparency and accountability requirements, and quantifying the scope of funding results.

Antonio Perianes-Rodríguez antonio.perianes@uc3m.es

Carlos Olmeda-Gómez carlos.olmeda@uc3m.es

Natalia R. Delbianco natalia.delbianco@unesp.br

Maria Cláudia Cabrini Grácio cabrini.gracio@unesp.br

- ¹ Department of Library and Information Science, Universidad Carlos III, C/Madrid, 128, 28903 Getafe, Spain
- ² Departamento de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Av. Hygino Muzzi Filho, 737 – Mirante, Marilia CEP 17.525-900, Brazil

Keywords Public funding · Linked open data · Co-funding networks

Introduction

Most governments plan how they intend to achieve their strategic objectives through research funding. Some of the typical objectives they seek are to obtain innovative research, get research results out into society, improve research capacity and skills, and use public resources efficiently (Lepori et al., 2023). Governments' missions include: to encourage the highest-quality research through competitive funding and to support frontier research across all fields, based on scientific excellence; to foster project-based research and stimulating innovation by promoting the emergence of collaborative multidisciplinary projects and encouraging collaboration between the public and private sectors; to promote the progress of science; and to advance national health, prosperity, and welfare by supporting basic research and people to create knowledge that transforms the future.

Also, ever more frequently agencies are explicitly requiring that the results of the funding they provide must be evaluated. Funders expect that their grants will help to bring about new, unpredictable scientific and technological discoveries; they wish to analyse trends in the research on offer and assess the impact of the funding they allocate to scientific output; they want to provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and technical resources and provide a source of information for policy formulation; and they need to monitor the activities they finance and gather the advice necessary to do a better job at designing and planning R&D policy actions and initiatives (AEI, 2022; ANR, 2022; ERC, 2022; NSF, 2022). However, questions remain about the strategies and instruments to use to evaluate and judge compliance with these objectives, because no such strategies or instruments have been defined.

Furthermore, although funders' functions also include disseminating and communicating the results of their activities and the results of the research they finance as an essential element of their accountability and transparency responsibility and obligations (Mugabushaka et al., 2022; Thelwall et al., 2023), this information is not often available. The reason may be the traditional intrinsic difficulty of obtaining pertinent, high-quality information of this kind plus the necessary metadata to accomplish the task successfully. The previous literature indicates some of the difficulties: linking the outputs associated with a specific grant or agency is a daunting prospect because of the lack of institutional instruments of explicit input-output data, and, even when such instruments exist, huge efforts are required to clean and normalise the data (Boyack et al., 2011); also, the data are usually given in the article acknowledgments, a non-structured section whose heterogeneous content may well include technical, ethics-related, and intellectual support in addition to financial assistance (Cronin et al., 1995). But the difficulties also depend on the variability of funding objectives and differences in expected results, so the job of measuring funding's effects on the scientific ecosystem is complicated even further. Some of the documented problems-such as the difficulty to assess concrete funding schemes because few authors specify which specific programme provided the support (Butler, 2001), and the fact that papers usually acknowledge funding from multiple agencies, thus making it difficult or problematic to measure the performance of a specific funder (Rigby, 2011)—have been overcome in recent years thanks to initiatives and sources such as those discussed below.

The organisations that support scientific research have a growing interest in evaluating their funding policies, an interest that has become a mission for them. Their efforts in this direction should focus on analysing past funding strategies in detail, to lay the foundations for better future decision making. Evaluation would guarantee that investments are put to good use and ensure funders' public accountability (MacLean et al., 1998). MacLean's 1998 paper was in fact one of the pioneers in the use of funding acknowledgements to measure the outputs and impact of funding in the context of the scientific environment. Three years earlier, Cronin said that comprehension of the social importance of FAs should be studied thoroughly to establish the advisability of including acknowledgements in the evaluation of scientific performance (Cronin et al., 1995). With this strategy, funders can determine how many publications they have funded and what visibility they have achieved, measuring what sum of their investments has been transformed into ground-breaking publications (Rigby, 2011). FAs help funders gain visibility and justify that the output of the research they fund responds to society's needs and objectives, increasing the transparency of their management (Álvarez-Bornstein et al., 2020). In short, the use of FAs makes it possible to analyse the relationships between funding input and research output, such as scientific publications (Grassano et al., 2017).

For all these reasons, funding agencies should keep databases of quality information about the financial assistance they grant and the resulting scientific output, to enable independent, reproducible analyses of their activity (Boyack et al., 2003). The infrastructure should offer open data with standardised identifiers to facilitate connection with other data sources that will expand and enrich the scope of the research results.

Publishers, too, share the responsibility of creating high-quality data infrastructure. Their tasks include gathering and sharing the funding data reported in funding acknowledgements. This important task is beneficial in two ways: it helps authors comply with their funding requirements by clearly and unequivocally indicating the results of the funding they have received, and it helps analyses evaluate the influence and effectiveness of funding policies (Cronin et al., 1995; Mugabushaka et al., 2022; Paul-Hus et al., 2016).

But FA-based evaluation is not perfect. It fails to illustrate the full complexity of the multiple perspectives and results stemming from scientific activity (mobility, patents, collaboration agreements, researcher training, contracts, events, equipment procurement, etc.). In addition, numerous technical drawbacks have been associated with FAs, such as data collection and processing errors, which have a negative impact on analysis quality and on the reliability and accuracy of results (Álvarez-Bornstein et al., 2017; Mugabushaka et al., 2022; Tang et al., 2017), and the omission of information due to voluntary or involuntary author errors (Grassano et al., 2017).

Consequently, this paper presents a methodology that uses open metadata sources to connect and analyse funded projects and the publications associated with them through author FAs. This methodology includes strategies for obtaining, linking, and analysing open data from freely accessible international sources. The method proposed aims to be a complementary tool for funding bodies to enhance the evaluation of their activities, especially those agencies with poor data infrastructures, and to extend the scope of their analyses. To verify the methodology's usefulness, the projects funded by the Spanish National Science and Research Agency (Agencia Estatal de Investigación, or AEI) from 2008 to 2020 are evaluated in order to test the robustness of the method using the AEI as a proof of concept. The study is divided into two tasks. The first is the analysis of funding information provided by the funder. The second is the analysis of the productivity and visibility of the publications that mention funding in their FAs, paying close attention to FA quality and the effects of co-funding on scientific influence, including analysis of the resulting co-funding networks. The methodological section traces in detail the steps necessary to implement the strategy for data acquisition and analysis.

The rest of the paper is organised as follows: Sect. "Objectives" states the paper's objectives; Sect. "Data and Methods" presents the data extraction methodology; Sect. "Results" reports the results of the analysis of projects and publications; Sect. "Discussion and Conclusions" summarises the main findings and conclusions; and Sect. "Limitations and Further Research" reflects on the methodology's limitations and presents future lines of research.

Objectives

This paper's main objective is to present a methodology using linked open metadata to diverse data sources to evaluate the performance of funding agencies.

The secondary objectives, which aim to test the method's usefulness and efficacy and achieve transparent, reproducible, independent evaluations, are the following:

- To ascertain the proportion of projects whose identifier is correctly given in the FAs of articles printed in scientific journals.
- To determine the scientific strategies of the National Plan through the project funding it provides, comparing this with the productivity and visibility of published outputs.
- To study the geographic implications of funding by examining the distribution of funds, productivity, and publication influence in Spain on a region-by-region basis.
- To conduct an institutional analysis, comparing institutions' ability to secure funding and their ability to produce highly visible publications.
- To evaluate co-funding's effects on the quantity and visibility of research output based on the total number of funders and their nationality, and to map the co-funding relationships the National Plan has created with leading international agencies.

Data and methods

To achieve this paper's main purpose, a connection had to be found between two databases that have no elements in common, the database containing the information on project funding and the database containing the publication metadata. It was therefore necessary, first, to use two intermediate sources to connect these two reciprocally incompatible datasets and, second, to broaden and extend the scope of the analyses of any of these four sources separately. This method highlights the important role analysts play in gathering and analysing the data offered by funding organisations to enhance the accuracy, solidity, and diversity of the results of quantitative studies, in this case quantitative studies based on FAs. The datasets that were combined (on citations, institutional analyses, funding analyses, geographic analyses, and relational analyses) seek to establish new, more-informative, more-robust methods of research suggested in the previous literature (Álvarez-Bornstein et al., 2020).

However, linking these sources is a far from trivial endeavour (Mugabushaka, 2020). Generating the final source with clean, normalised information on articles and projects requires hard, tedious disambiguation and data-cleaning work to harmonise the sources (Boyack, 2009; Boyack et al., 2011). Nor is the linking procedure free of problems and errors, as will be shown below.

Source	Metadata
Funder Registry	DOI, funder DOI, publication title, funder name, and funder award
Spanish National Science and Research Agency	project reference, organisation, centre, total funding, year, tax identification number, and project title. The subject area descrip- tor was extracted from the project reference

Table 1 Sources and list of metadata downloaded

The data on funded domestic projects were downloaded from the AEI portal in mid-February 2022. At that time the search engine referring to granted funding was not working, and the data were downloaded from decisions published in.pdf file form. Two tables were extracted from each document: a table of grant data and institutional affiliation and a table of economic data on the assistance granted to each project. The data available in the. pdf files did not match the data furnished by the search engine (the search engine yielded a smaller number of metadata). Thus, 38,384 projects were found in the research calls shown in Table 1, for 2008 to 2020. The names of the calls were changed in 2018, although the change did not affect the aggregate analysis of projects and publications.

The foremost of the problems detected in the information downloaded from the AEI were the lack of homogeneity in the names given to identify funding recipients (i.e., numerous institutions appeared under several variants of the same name) and the absence of data about the recipient centres (more than 1000 grants awarded to the Spanish National Research Council (Consejo Superior de Investigaciones Científicas, or CSIC) fail to identify the particular centre concerned). Not only did organisations and centres have to be subjected to disambiguation, but the identifier assigned to the funded institutions by the Research Organization Registry (ROR) had to be included as well (for 424 of the 478 institutions). ROR is a community-led registry of open, sustainable, usable, unique identifiers for every research organisation in the world (ROR, 2022).

The data on publications with funding information were obtained in mid-January 2022 with the API of the Funder Registry (FR) managed by Crossref. Eight variants of the names of the ministries involved in project funding under the National Plan were used. A total of 105,683 distinct publications were downloaded from the FR, 104,891 of which were available in OpenAlex (Priem et al., 2022). In this case, the "funder award" variable was subjected to a painstaking cleaning process, so that this dataset could be connected with the AEI dataset. Semiautomatic disambiguation of the "funder name" and "funder country" variables in the 489,604 rows (multiplicative counting) was also performed, to yield robust results for the analysis of international co-funding networks. The downloaded metadata are summarised in Table 1.

The problems encountered in the Funder Registry data are identical to the problems indicated in previous papers, and they can be traced to errors in the automatic information extraction tools that require separating multiple funders from within the funding organisation field, merging variations in spelling or abbreviations of the same funder, isolating unintelligible funder names, and finally categorising identified funding organisations (Gok et al., 2015). The reliability and correctness of the information furnished by Crossref is evaluated in recent papers (Kramer et al., 2022; Mugabushaka, 2022; Mugabushaka et al., 2022).

Finally, the clean data linked to projects, funded publications, and institutions were connected with OpenAlex (version November 2022, downloaded in January 2023) to add the citation data to the publications downloaded from the Funder Registry.



Fig. 1 Linked open data. Schema of sources and normalised identifiers used

The final result enabled 58,471 distinct publications (55.7% of the total number of publications downloaded) mentioning 17,547 projects (45.5% of the total number of projects downloaded) to be connected. Figure 1 charts the linking procedure and the normalised identifiers used to connect the four data sources.

The indicators used in this paper are:

- (a) P: total number of projects.
- (b) PPubs: total number of projects acknowledged in publications.
- (c) Pubs and PrPubs: total number of publications acknowledging funding and proportion of publications acknowledging funding.
- (d) PTop1: proportion of publications acknowledging funding that are in the first percentile and thus are considered highly cited publications. The proportion of publications in Top 1% indicator is the top one percent most cited publications in a given subject category, year and document type divided by the total number of publications in a given set of documents, displayed as a percentage (Clarivate, 2018; Roberge, 2021). It is important to highlight that the Top 1% indicator has been applied to the OpenAlex dataset. The categories provided by Scopus (ASJC 2-digit and 4-digit code) were linked to the OpenAlex-listed publications using ISSN numbers. Multi-categorised publications have as many percentiles as categories.
- (e) PUncited: proportion of publications acknowledging funding that are never cited.
- (f) Number of funders: number of funding agencies per publication.
- (g) TF: total funding (in euros).
- (h) TF/P: average amount of funding per project.
- (i) PPubs/P: percentage of projects acknowledged in publications.
- (j) Pubs/PPubs: average number of publications per project acknowledged in publications.

Tables were created grouping Spanish calls, regions, and institutions by proximity in terms of these indicators. The grouping was done by means of cluster analysis, using Ward's method, with squared Euclidean distance measurements and normalised indicators. Cluster analysis is a statistical multivariate method to classify objects of a universe. This method groups objects according to a set of variables based on a geometric similarity measure. Thus, the objects of a cluster are more similar to each other (intragroup) than to those in other clusters (intergroup) based on the variables used (Hair et al., 2005; Reis, 2001). Cluster analysis was used to maximise the intergroup intragroup differences with respect to the intragroup variations. Hierarchical agglomerative clustering was applied because it is commonly present in statistical packages as SPSS. Ward linkage was selected from those more usual in hierarchical clustering such as single-linkage, complete-linkage, weighted average linkage, or centroid linkage, among others. Squared Euclidean distance, a measure based on the smallest distance among pairs of points from the objects, was used. Clustering of calls, regions and Spanish institutions took into consideration the nine scientometric indicators shown on Tables 2, 3, 4. A normalisation procedure was applied using z-score due to the scale differences of the variables analysed. This process normalises every value in a dataset such that the mean of all the values is 0 and the standard deviation is 1 (Hair et al., 2005; Reis, 2001).

The difference between the clusters generated for each variable (calls, regions, and institutions) was evaluated using the nonparametric Kruskal–Wallis test (or one-way ANOVA on ranks). This tests the null hypothesis that three or more groups have the same distribution assuming the ordinal scale of the analysed variable. It is an alternative to the parametric ANOVA in case of small samples or samples that do not meet the ANOVA requirements (Vieira, 2003). The test was also applied to the nine indicators shown on Table 2, 3, 4, using the indicators as dependent variables. The statistical analyses were performed with SPSS software, version 22. The level of significance was 0.05.

VOSviewer software was used to create the co-funding networks based on fractional counting (Perianes-Rodriguez et al., 2016). Additional socio-economic indicators were obtained from the Spanish National Statistics Institute website. All the datasets used in this study are available at (https://zenodo.org/record/7767579).

Results

This section addresses some of the most significant aspects of the characterisation of Spanish funding through Spanish national plans, fundraising ability according to region and institution, and the productivity and visibility of the output of funded research. The results presented below are only a sample of the breadth and depth of the analyses that can be carried out with the open metadata used in this study, and they help to exemplify the use and usefulness of linked public data in evaluating science and technology funding.

Calls

Table 2 presents the proportional distribution of total funding (4,470,557,183 euros) according to call. The calls are defined by the AEI on the basis of an ad-hoc list dividing research activity into what are called "subject area descriptors".¹ The data indicate that the largest budget is for R&D projects (17.3%), Biomedicine (8.8%), and Fundamental Biology (7.7%). The calls with the smallest budgets are Transport (0.5%), Sport Sciences (0.3%), and Gender Studies (0.1%). Curiously, the calls receiving the lowest proportions of AEI

¹ https://tinyurl.com/2wubj9u8 [Accessed: 20/07/2023].

Call	Abbr	% Funding	% Projects with publications
R&D projects	PID	17.3	29.9
Biomedicine	SAF	8.8	59.3
Fundamental Biology	BFU	7.7	48.2
Agrofood Technology	AGL	6.3	57.3
Chemistry	CTQ	5.6	66.4
Research Challenges	RTI	5.6	66.8
Biodiversity, Earth Sciences and Climate Change	CGL	5.5	57.6
Material Sciences	MAT	4.1	66.0
Communication Technologies	TEC	3.4	56.4
Biotechnology	BIO	3.2	54.4
Astronomy and Astrophysics	AYA	3.2	38.2
Computer Sciences	TIN	2.6	52.4
Environmental Sciences	CTM	2.6	60.8
High Energy Physics	FPA	2.5	48.2
Knowledge Creation	PGC	2.4	58.0
Industrial Design and Production	DPI	2.4	34.1
Energy	ENE	1.7	52.1
Art and History	HAR	1.7	15.5
Philology and Philosophy	FFI	1.7	12.8
Physics	FIS	1.7	70.9
Space Sciences	ESP	1.6	52.8
Mathematics	MTM	1.5	65.5
Psychology	PSI	1.3	62.0
Social Sciences	CSO	1.2	28.2
Economy	ECO	1.1	54.7
Construction	BIA	0.9	49.5
Law	DER	0.9	3.2
Education	EDU	0.6	34.8
Transport	TRA	0.5	39.5
Sport Sciences	DEP	0.3	55.7
Gender Studies	FEM	0.1	23.2

Table 2 Proportion of funding by call

funding are not the calls with the lowest proportion of funded projects acknowledged in publications.

Table 3 presents the six call clusters (C1 to C6), which are grouped according to the indicators for projects, publications, funding, and highly cited or uncited publications, including their averages and coefficients of variation (CVs). The table also furnishes general descriptive statistics about the calls as a set.

The average percentage of projects mentioned in publication FAs for the entire set of 31 calls is 47%, with figures ranging from 71% for Physics to 3% for Law. In all the areas in the humanities, less than 15% of projects are mentioned in FAs. The proportion for highly

call
by
funding
and
publications,
projects,
of
Clusters
Table 3

	Code	Call	Р	Ppubs	Pubs	TF	Ptop1	Puncited	TF/P	Ppubs/P (%)	Pubs/PPubs
C1 N = 4	AGL	Agrofood Technol.	2143	1229	3494	28,21,39,918	1.13	6.84	1,31,657	57	2.82
	CGL	Earth Sciences	2084	1183	3410	24,72,03,983	0.94	9.50	1,18,620	57	2.88
	SAF	Biomedicine	2043	1203	3558	39,56,20,094	1.57	7.03	1,93,647	59	2.96
	BFU	Fundamental Biology	1742	833	1840	34,33,93,564	1.09	6.90	1,97,126	48	2.21
		Average	2003	1110	3064	31,70,89,390	1.18	7.60	160	55	2.72
		CV	%6	17%	27%	21%	23%	17%	26%	6	13%
C2 N = 6	MTM	Mathematics	1121	732	4170	6,87,37,523	0.22	20.41	61,318	65	5.70
	PGC	Knowledge Creation	1135	626	2724	10,80,61,755	0.66	33.99	95,209	55	4.35
	RTI	Research Challenges	1853	1177	3824	25,13,87,785	0.68	32.92	1,35,665	64	3.25
	СТQ	Chemistry	1917	1273	6,966	25,17,91,422	0.72	5.46	1,31,347	66	5.47
	MAT	Material Sciences	1493	983	4,863	18,25,46,474	1.13	6.97	1,22,268	66	4.95
	FIS	Physics	784	556	4,441	7,51,38,446	0.88	13.96	95,840	71	7.99
		Average	1384	891	4498	15,62,77,234	0.72	19.00	1,12,917	65	5.28
		CV	32%	33%	31%	54%	42%	66%	26%	8	30%
C3 N = 6	CSO	Social Sciences	1013	278	478	5,20,99,349	0.21	22.38	51,431	27	1.72
	HAR	Art and History	1932	288	546	7,79,36,936	0.18	24.91	40,340	15	1.90
	DER	Law	1233	38	54	3,94,56,284	0.00	31.48	32,000	3	1.42
	FEM	Gender Studies	66	23	44	29,51,165	0.00	29.55	29,810	23	1.91
	EDU	Education	563	192	374	2,48,00,103	2.14	20.86	44,050	34	1.95
	FFI	Philosophy and Language	1932	222	398	7,70,32,912	1.26	34.92	39,872	11	1.79
		Average	1129	174	316	45,712,792	0.63	27.35	39,584	19	1.78
		CV	65%	67%	68%	65%	139%	20%	20%	60	11%
C4 N = 4	AYA	Astronomy and Astrophysics	521	198	730	14,08,32,711	2.33	16.44	2,70,312	38	3.69
	BIO	Biotecnology	708	382	1087	14,23,94,131	2.21	6.26	2,01,122	54	2.85
	FPA	High Energy Physics	388	187	1137	11,13,49,117	2.02	8.62	2,86,982	48	6.08
	ESP	Space Sciences	195	101	379	7,13,78,734	1.06	17.68	3,66,045	52	3.75
		Average	453	217	833	11,64,88,673	1.91	12.25	2,81,115	48	4.09
		CV	48%	54%	42%	29%	30%	46%	24%	15	34%

Table 3 (continued)

	Code	Call	Р	Ppubs	Pubs	TF	Ptop1	Puncited	TF/P	Ppubs/P (%)	Pubs/PPubs
C5 N = 10	BIA	Construction	382	185	434	4,02,69,766	0.69	8.76	1,05,418	48	2.35
	TRA	Transport	210	83	213	2,04,11,592	0.47	9.39	97,198	40	2.57
	DEP	Sport Sciences	149	81	184	1,13,63,442	1.09	14.67	76,265	54	2.27
	DPI	Industrial Design	915	307	660	10,65,20,205	0.61	15.30	1,16,416	34	2.15
	ECO	Economics	1054	574	1707	4,82,79,509	1.41	17.69	45,806	54	2.97
	ISd	Psychology	936	573	1554	5,93,56,824	0.97	14.03	63,415	61	2.71
	CTM	Environmental Sciences	800	481	1352	11,44,48,157	0.52	7.62	1,43,060	60	2.81
	ENE	Energy	633	327	1089	7,80,47,372	1.10	9.27	1,23,298	52	3.33
	TEC	Communication Technol.	1105	619	2464	15,36,97,564	0.85	11.81	1,39,093	56	3.98
	NIL	Computer Sciences	1183	605	2299	11,72,69,627	1.91	12.74	99,129	51	3.80
		Average	737	384	1196	7,49,66,406	0.93	12.13	1,00,910	51	2.89
		CV	51%	56%	%69	62%	41%	27%	31%	17	22%
C6 N = 1	PID	R&D Projects	6118	1419	3215	77,46,40,719	0.56	54.12	1,26,617	23	2.27
General $N = 31$		Minimum	66	23	44	29,51,165	0.00	5.46	29,810	3	1.42
		Maximum	6118	1419	6966	77,46,40,719	2.33	54.12	3,66,045	71	7.99
		Average	1238	547	1924	14,42,11,522	0.98	17.18	1,21,948	47	3.25
		CV	%68	%LL	%06	106%	64%	66%	65%	38	46%
- - -		-									

Bold numbers just represent average values for every single cluster

P projects. Pubs publications. PPubs projects acknowledged in publications. TF total funding. PTop1 proportion of highly cited publications. PUncited proportion of uncited publications. TF/P average funding per project. PPubs/P proportion of projects acknowledged in publications. Pubs/PPubs average of publications per project acknowledged in publications.

Table 4 Projects,	publications, and funding by	Spanish re	egion							
	Region	Р	Ppubs	Pubs	TF	PTop1	Puncited (%)	TF/P	PPubs/P (%)	Pubs/PPubs
C1 N = 7	Asturias	720	342	1140	7,40,30,309	96.0	15	1,02,820	48	3.30
	Canary Islands	714	325	1032	7,68,55,520	0.97	17	1,07,641	46	3.20
	Murcia	811	375	1338	8,35,88,527	1.12	17	1,03,068	46	3.60
	Aragon	1266	611	2195	14,38,85,422	0.73	15	1,13,654	48	3.60
	Castile-La Mancha	584	271	891	6,10,81,577	0.56	17	1,04,592	46	3.30
	Cantabria	502	215	683	6,30,17,295	0.88	19	1,25,532	43	3.20
	Navarre	581	252	817	5,56,49,042	1.96	19	95,781	43	3.20
	Average	740	342	1157	7,97,29,670	1.03	17	1,07,584	46	3.34
	CV	34%	38%	44%	38%	44%	6	%6	4	5%
C2 N = 6	Extremadura	329	127	486	2,86,36,961	0.82	18	87,042	39	3.80
	La Rioja	152	64	257	1,25,18,836	0.39	18	82,361	42	4.00
	Castile and Leon	1532	632	2232	14,17,24,341	0.85	16	92,509	41	3.50
	Galicia	1827	LLL	2577	18,27,40,270	1.01	15	1,00,022	43	3.30
	Basque Country	1519	667	2749	15,14,54,495	0.91	16	707,66	44	4.10
	Balearic Islands	416	203	926	4,00,42,175	0.65	14	96,255	49	4.60
	Average	963	412	1538	9,28,52,846	0.77	16	92,983	43	3.90
	CV	$^{\% LL}$	76%	72%	80%	29%	6	8%	8	11%
C3 N = 2	Catalonia	8171	3595	12,404	1,00,39,69,174	1.17	15	1,22,870	4	3.50
	Madrid	9846	4199	14,104	1,29,91,85,316	0.91	15	1,31,951	43	3.40
	Average	6006	3897	13,254	1,15,15,77,245	1.04	15	1,27,410	43	3.40
	CV	13%	11%	9%6	18%	18%	4	5%	2	2%
C4 N = 2	Andalusia	5338	2407	7689	55,95,35,722	0.86	16	1,04,821	45	3.20
	Comunidad Valenciana	3690	1757	6852	42,75,08,933	0.74	15	1,15,856	48	3.90
	Average	4514	2082	7271	49,35,22,328	0.8	16	1,10,339	46	3.55
	CV	26%	22%	8%	19%	11%	4	7%	4	14%

_
(panu
conti
4
ble
Tal

	Region	Р	Ppubs	Pubs	TF	PTop1	Puncited (%)	TF/P	PPubs/P (%)	Pubs/PPubs
General $N = 17$	Minimum	152	64	257	1,25,18,836	0.39	14	82,361	39	3.20
	Maximum	9846	4,199	14,104	1, 29, 91, 85, 328	1.96	19	1,31,951	49	4.60
	Average	2235	989	3434	25,91,42,583	0.91	16	1,05,087	45	3.60
	CV	129%	127%	124%	142%	36%	6	13%	6	11%
Bold numbers inst	ramecant original for	ouio mane a	de chieter							

angine **1146** rdat tenf eta

P projects. *Pubs* publications. *PPubs* projects acknowledged in publications. *TF* total funding. *PTop1* proportion of highly cited publications. *PUncited* proportion of uncited publications. *TF/P* average funding per project. *PPubs/P* proportion of projects acknowledged in publications. *Pubs/PPubs* average of publications per project with publications tions.

cited papers averages out to 0.98%, that is, slightly lower than the expected 1%. By calls, Astronomy and Astrophysics projects are the most influential (2.3%). Gender Studies and Law, on the other hand, have no cited papers at all. The average percentage of uncited publications is 17.18%. The data show that the highest budget is for R&D projects, and the call with the lowest budget is Gender Studies. However, when these figures are relativised according to the total number of projects, the highest average budget is for the Space Sciences call.

The first cluster, C1, contains four calls, Agrofood Technology, Earth Sciences, Biomedicine, and Fundamental Biology. In terms of CV, this is the most homogeneous group the cluster analysis produced (CV = 9-27%), which indicates the representativeness of the averages of these indicators. The calls in this group are characterised by a large number of funded projects with high total funding (TF) figures and a high percentage of projects acknowledged in publications that mention the AEI's funding in their FAs (48 to 59%). In terms of bibliometric indicators, this group's calls are observed to usually have an average of 2.7 publications per funded project and a low proportion of uncited publications. Around 1.2% of its publications are included in the top 1%.

Group C2 is made up of six calls, Mathematics, Knowledge Creation, Research Challenges, Chemistry, Material Sciences, and Physics. It is the group that produces the highest number of research publications, an average of 5.28 papers (CV = 30%). On average, 65% of the funded projects are associated with publications. This group tends to have one of the lowest proportions of publications in the high-impact PTop1 category (0.72% on average, CV = 42%). The proportion of uncited publications ranges widely in this cluster. This group contains Chemistry, the call with the highest number of publications (Pubs), and Physics, the call with the highest proportion of projects acknowledged in publications that mention funding in their FAs and the highest number of publications per project.

Cluster C3 is also made up of six calls, Social Sciences, Art and History, Law, Gender Studies, Education, and Philology and Philosophy. Its leading characteristic is that it has the lowest percentage of projects acknowledged in publications acknowledging funding in their FAs. In addition, C3's projects have a low publication productivity level, producing an average of 1.78 publications per project (CV = 11%), a high percentage of uncited publications (only cluster C6 has a higher percentage), and the lowest investment per project (average TF/P=39.6). Furthermore, C3 is a highly heterogeneous group in terms of PTop1; the cluster includes Education, the call with the third-highest proportion of articles in the top 1%, but also the two calls that have no articles are all in the top 1% (Law and Gender Studies).

Cluster C4 is made up of the calls for Astronomy and Astrophysics, Biotechnology, High Energy Physics, and Space Sciences. Its calls have the highest TF/P, the highest percentage of publications in the top 1% (with the exception of the Space Sciences call, whose proportion is only slightly above the general trend), and a high number of publications mentioning funded projects in their FAs. Space Sciences is also the call with the greatest funding per project (TF/P), and the Astronomy and Astrophysics call presents the highest percentage of highly cited publications.

C5 is the largest cluster. It contains 10 calls. Its most homogeneous characteristics are the fact that it has an average of around 2.9 publications mentioning funding in their FAs and a low percentage of uncited articles.

Cluster C6 contains only the R&D call, which outperforms all the other calls in terms of five of the indicators (P, PPubs, TF, PPubs/P, and PUncited).

Application of the Kruskal–Wallis test and the post-hoc difference test revealed the following similarities and differences among the clusters for the nine indicators. In total

projects, the C4 calls tend to resemble the C5 calls, and the C3 calls tend to resemble the C2 and C1 calls, while C6 is completely different from the rest. In terms of the PPubs/P indicator, C3, C4, and C5 have similar tendencies, and C1 is close to C2. Again, C6 lies very far from the other clusters.

Only clusters C3 and C4 have similar trends in regard to the Pubs bibliometric indicator; the rest differ from both C3/C4 and from each other. For the TF indicator, C6 has far larger figures than the other clusters. C1 is also significantly different from the rest.

In terms of the PTop1 indicator, C2 is similar to C3, tending to display lower proportions of high-visibility papers. C4, on the other hand, is different from all the other clusters, with higher proportions of highly cited articles. Clusters C1 and C5 resemble each other in this indicator.

Clusters C1, C4, and C5 tend to resemble each other in terms of PUncited. Cluster C3 stands apart from all the rest, with a high proportion of uncited papers.

Clusters C1, C2, and C5 present similar TF/P indicator behaviour. Cluster C3 tends to have the lowest figures, and C4, the highest figures for this indicator.

The indicators PPubs and Pubs/PPubs present a similar picture for clusters C1, C4, and C5. C2, however, tends to have the highest figures for these indicators, and C3, the lowest.

Regions

Table 4 displays the data by regions, grouped into four clusters. Four regions of Spain (Andalusia and Valencia in C4, Catalonia and Madrid in C3) account for 71% of the projects and publications and nearly 75% of the funding. Of these four, only Catalonia displays a proportion of highly cited publication higher than the expected 1%. The average percentage of highly cited publications is 0.91%, and the average percentage of uncited papers is very similar to the figure found in the analysis of calls (16.3%).

PUncited does not vary greatly by region. For the set of 17 regions, the variation was small (CV=9%), indicating that the proportion of uncited publications (around 16%) is very similar in all regions. The proportion of projects acknowledged in publications (PPubs/P) also displays little inter-regional variability (CV=6%), which shows that on average 45% of funded projects produce publications. On the other hand, P, PPubs, Pubs, TF, and PTop1 are highly disperse.

Cluster C1 is made up of four regions. With the exception of Aragon, it is a very homogeneous group in terms of its low number of funded projects (average = 740, CV = 34%) and its low number of projects acknowledged in publications (average = 342, CV = 38%).

Cluster C2 covers six regions. It is the group with the highest internal variability in all indicators. This group contains the minimum figure for all indicators except Pubs/PPubs. Furthermore, this group also features the highest figure for Pubs/PPubs and PPubs/P.

C3, a two-region cluster, contains Madrid, the region with the highest figures in five of the indicators (P, PPubs, Pubs, TF, and TF/P). C2 is the group with the lowest internal variability in all indicators. Its average is much higher than that of the other clusters in indicators P, PPubs, Pubs, TF, and TF/P.

Cluster C4, which is made up of two regions, also presents high internal homogeneity. Its indicators on average reach figures much higher than those of C1 and C2, yet lower than those of C3, with the exception of PPubs/P and P/PPubs.

Figure 2 presents a geographic view of the distribution of funding by regions in terms of percentages.



Fig. 2 Funding proportion by region

The regions in blue (Madrid and Catalonia) receive over half the funding (52%). La Rioja, on the contrary, draws only 0.3% of the funding. Cantabria, Castile-La Mancha, Extremadura, the Balearic Islands, and Navarre together hold approximately 5.6% of the funding granted by the AEI.

Institutions

Four hundred twenty one institutions received funding. Of them, 347 reported publications. Table 5 contains the 16 organisations with over 600 funded projects apiece (45.2% of all projects and 40.6% of all funding). The institutions are grouped into six clusters according to the similarities revealed by indicator analysis. Only four of these institutions present a proportion of highly cited publications that is higher than the expected 1% (the sole institution in C6, the two institutions in C3, and one of the institutions in C2). The average proportion of uncited papers (16.2%) is similar to that found in the analyses of calls and regions.

Initially, clusters C1, C2, C3, and C5 behave very similarly in terms of the Pubs/PPubs indicator, with an average of three to four publications per project and a CV of under 13%. This result concurs with the fact that the universities in these clusters have similar percentages of projects acknowledged in publications (PPubs/P), with averages ranging between 39 and 44%. On the other hand, C4 tends to have more publications associated with projects (an average of 4.2), possibly because it tends to have more projects acknowledged

 Table 5
 Clusters of projects, publications, and funding by institutions (>600 funded projects)

		•								
	Institution	Ρ	Ppubs	Pubs	TF	PTop1	PUncited	TF/P	PPubs/P (%)	Pubs/PPubs
C1 N = 3	Univ. Sevilla	1083	487	1624	10,68,26,579	0.37	17.36	98,640	45	3.30
	Univ. Zaragoza	930	423	1328	9,39,93,078	0.53	18.90	1,01,068	45	3.10
	Univ. Autonoma Barcelona	1240	516	1877	11,61,56,007	0.53	18.01	93,674	42	3.60
	Average	1084	475	1610	10,56,58,555	0.48	18.10	97,794	44	3.40
	CV	14%	10%	17%	11%	19%	4%	4%	5	7%
C2 N = 4	Univ. Santiago de Compostela	872	355	1189	8,87,31,322	1.01	15.90	1,01,756	41	3.30
	Univ. Oviedo	627	293	976	6,11,67,152	0.72	16.09	97,555	47	3.30
	Univ. Politecnica Madrid	736	337	876	8,76,72,438	0.34	17.12	1,19,120	46	2.60
	Univ. Carlos III de Madrid	639	255	894	5,31,38,012	0.45	22.48	83,158	40	3.50
	Average	719	310	984	7,26,77,231	0.63	17.90	1,00,397	43	3.20
	CV	16%	15%	15%	25%	48%	17%	15%	8	13%
C3 N = 2	Univ. Granada	1289	550	1993	10,68,18,679	1.05	15.76	82,869	43	3.60
	Univ. Autonoma Madrid	1149	493	1907	11,82,63,639	1.15	14.47	1,02,927	43	3.90
	Average	1219	522	1950	11,25,41,159	1.10	15.12	92,898	43	3.70
	CV	8%	8%	3%	7%	6%	6%	15%	0.4	5%
C4 N = 4	Univ. Valencia	1183	524	2272	12,48,67,691	0.44	15.10	1,05,552	44	4.30
	Univ. Pais Vasco	1041	472	1995	10,04,05,263	0.85	15.39	96,451	45	4.20
	Univ. Politecnica Cataluña	903	455	2025	11,35,82,490	0.84	17.48	1,25,783	50	4.50
	Univ. Politecnica Valencia	842	424	1637	9,28,28,945	0.49	16.92	1,10,248	50	3.90
	Average	992	469	1,982	10, 79, 21, 097	0.66	16.22	1,09,509	48	4.20
	CV	15%	9%6	13%	13%	34%	7%	11%	7	6%
C5 N = 2	Univ. Barcelona	1914	756	2723	19,70,58,753	0.95	13.48	1,02,957	39	3.60
	Univ. Complutense	1770	691	2735	17,14,81,257	0.62	16.67	96,882	39	4.00
	Average	1842	724	2729	18,42,70,005	0.79	15.08	916,919	39	3.80
	CV	6%	6%	0.3%	10%	30%	15%	4%	0.8	7%
C6 N = 1	CSIC (*)	1133	403	1,162	18,23,70,525	1.03	7.40	1,60,963	36	2.90

(continued)	
Table 5	

	Institution	Ь	Ppubs	Pubs	TF	PTop1	PUncited	TF/P	PPubs/P (%)	Pubs/PPubs
General $N = 16$	Minimum	627	255	876	5,31,38,012	0.34	7.40	82,869	36	2.60
	Maximum	1914	756	2735	19,70,58,753	1.15	22.48	1,60,963	50	4.50
	Average	1084	465	1701	11,34,60,114	0.71	16.16	1,04,975	43	3.60
	CV	33%	28%	35%	35%	38%	19%	18%	6	14%
Bold numbers just	represent average values for every s	single clus	ter							

b ů, 5 B ا

P projects. Pubs publications. PPubs projects acknowledged in publications. TF total funding. PTop1 proportion of top 1% publications. PUncited proportion of uncited pub-lications. TF/P average funding per project. PPubs/P proportion of projects acknowledged in publications. Pubs/PPubs average of publications per project with publications *Documents without a specific centre in publications (PPubs/P=48%). Cluster C6 has the lowest PPubs/P and also the lowest average number of publications per project. Strikingly, while all the institutions in the other clusters are universities, the only institution in cluster C6 is not a university.

The first cluster (C1) is made up of three universities. This group is quite homogeneous, since the CVs of all its indicators are under 20%. Although the institution with the highest proportion of uncited publications (PUncited = 22.4) does not belong to C1, C1 is nonetheless the group with the highest average number of uncited publications (18.1%) and low behaviour variability (CV=4%). It is also the group with the lowest average number of publications among the most-visible publications (average PTop1=0.5; CV=19%).

Cluster C2 consists of four universities. Compared with the other clusters, this group presents the lowest averages for total projects (P), projects acknowledged in publications (PPubs), total publications (Pubs), and Total Funding (TF). It also has the second-highest average for uncited publications and the second-lowest average for highly cited publications (PTop1). C2 contains the institutions with the lowest figures for six of the indicators (P, PPubs, Pubs, TF, PTop1, and Pubs/PPubs). Furthermore, it has the institution with the highest PUncited (22.5%) as well.

The two-university C3 cluster is the group with the most-homogeneous indicator figures, since its CV ranges between 0.4 and 15%. Although it is the cluster with the second-highest number of projects (average=1.2 and CV=8%), it is the group with the lowest average figures for funding per project. Both universities in this cluster have slightly more than 1% of their publications in the top 1%, and their figures for this indicator are in fact that highest of all 16 institutions analysed.

Group C4 consists of four universities. Its main characteristic has to do with its bibliometric indicators, since it presents the highest percentage of projects acknowledged in publications and the highest average number of publications per project. However, its universities number among those with the lowest numbers of AEI-funded projects. The institutions with the highest figures for the indicators PPubs/P and Pubs/PPubs are in C4.

Made up of two universities, C5 is the cluster with the most similar figures for all indicators analysed except PTop1. This cluster contains the universities with the highest figures in four of the indicators analysed (P, PPubs, Pubs, and TF). Thus, a certain duality is observed with cluster C2.

Cluster C6 contains a single institution. It has the lowest figures in two indicators, PPubs/P and PUncited. Furthermore, it has the highest average amount of funding per project.

Table 6 contains a list of the institutions with the highest proportions of highly cited papers, all of which have a visibility of far over the expected 1%. The characteristics common to these organisations are that each is a public research centre, each is highly specialised in its scientific field, each has a recognised national and international reputation, and each has a proportion of no-visibility papers that is lower than the general average.

Projects

Table 7 presents the projects with the highest numbers of mentions in FAs. The projects with the highest productivity are predominantly in Physics, Mathematics, and Chemistry. The proportion of highly cited papers in these projects is either very high or zero. Interestingly, some high-productivity projects present a high proportion of uncited publications; this may be explained by the great immediacy of the publications and the lower obsolescence of the specialities in question (Mathematics or Economics).

Institution	Р	Pubs	TF	PTop1	PUncited
Centro de Investigacion Ecologia y Aplicaciones Forestales (CREAF)	44	56	7,444,265	7.1	0.0
Centro de Edafologia y Biologia Aplicada del Segura (CEBAS)	76	119	12,494,960	6.7	0.8
Hospital Vall d'Hebron	38	45	6,939,060	6.7	6.7
Centre de Regulacio Genomica (CRG)	127	111	36,723,426	6.3	2.7
Instituto de Catalisis y Petroquimica (ICP)	50	162	8,613,163	5.6	1.9
Centro Nacional de Investigaciones Cardiovasculares (CNIC)	99	90	27,707,901	5.6	1.1
Centro Nacional de Investigacion sobre Evolucion Humana (CENIEH)	15	55	1,874,491	5.5	5.5
Institut de Recerca de l'Energia Catalunya (IREC)	26	41	4,091,492	4.9	4.9
Institut de Recerca Biomedica (IRB)	100	127	32,843,552	4.7	3.1
FISABIO Generalitat Valenciana	18	43	3,145,298	4.7	7.0
Instituto de Productos Lacteos Asturias (IPLA)	19	43	3,358,976	4.7	2.3
Centro Nacional de Investigaciones Oncologicas (CNIO)	88	67	26,831,054	4.5	4.5

 Table 6
 Projects, publications, and funding in highly cited institutions (at least 40 pub)

P projects. Pubs publications. PPubs projects acknowledged in publications. TF total funding. PTop1 proportion of top 1% publications. PUncited proportion of uncited publications

Table 7Projects, publications,and funding by projects (> 50pubs)	Р	Pubs	TF	PTop1	PUncited
	CTQ-2017-85821-R	133	105,270	0.00	9.02
	MTM-2016-77278-P	110	140,118	0.00	39.09
	FIS-2016-76363-P	70	119,790	4.29	7.14
	FIS-2015-67161-P	69	237,160	1.45	15.94
	FIS-2017-87519-P	64	169,400	1.56	18.75
	TIN-2015-65316-P	62	832,722	0.00	22.58
	FIS-2016-76617-P	60	192,995	0.00	8.33
	CTQ-2014-57393-C2-1-P	58	108,900	0.00	0.00
	MTM-2014-57129-C2-1-P	55	102,971	0.00	27.27
	CTQ-2015-63997-C2-2-P	54	45,012	0.00	1.85
	FIS-2016-75862-P	54	163,350	7.41	14.81

P projects. Pubs publications. TF total funding. PTop1 proportion of highly cited publications. PUncited proportion of uncited publications

The highest-visibility projects are predominantly in High Energy Physics, Chemistry, Agrofood Technology and Material Sciences, but there are also highly influential projects in Environmental Sciences, Biomedicine or Physics (Table 8).

International Co-funding networks

Table 9 presents the data on co-funding for publications that are funded under the Spanish National Plan. The list of collaborating countries includes the countries participating in at

Р	Pubs	TF	PTop1	PUncited
FPA-2013-41042-P	11	1,863,400) 36.4	36.4
CTQ-2013-41507-R	13	1,950,520	0 30.8	0.0
AGL-2010-22319-C03-01	15	2,087,250	0 26.7	0.0
MAT-2014-59134-R	13	1,519,518	8 23.1	0.0
DEP-2013-47540-R	10	1,452,000	0 20.0	0.0
ENE-2013-44330-R	10	2,050,000	0 20.0	0.0
FFI-2016-76799-P	10	556,600	20.0	0.0
SAF-2017-88276-R	10	2,420,000	0 20.0	50.0
CTM-2014-58481-R	11	2,835,030	0 18.2	0.0
FIS-2016-78904-C3-2-P	11	598,950	18.2	18.2

P projects. *Pubs* publications. *TF* total funding. *PTop1* proportion of highly cited publications. *PUncited* proportion of uncited publications

Country	Pubs	PrPubs	PTop1
European Union	17,644	42.6	30.4
USA	4064	9.8	12.7
United Kingdom	1666	4.0	6.6
Germany	1812	4.4	5.1
France	1530	3.7	5.1
China	1153	2.8	3.4
Switzerland	581	1.4	2.7
Australia	416	1.0	2.4
Brazil	1092	2.6	2.2
Portugal	1062	2.6	2.0
Italy	844	2.0	1.8
Belgium	445	1.1	1.8
Netherlands	325	0.8	1.4
Japan	561	1.4	1.3

Pubs publications. *PrPubs* proportion of publications. *PTop1* proportion of highly cited publications

least 19 highly cited publications. European funding is drawn from both research funds and structural funds.

Figure 3 shows the countries' differences in proportion in terms of their total number of publications and their number of highly cited publications (according to the data in the table above). A considerable increase in international representativeness can be seen in the highly cited publications.

Figure 4 presents the proportion of highly cited publications and the number of funding countries. Figure 5 then compares the proportion of publications in the top 1% and the number of funders. In both cases, it is clear that, as publications gain more funding agencies, they gain higher visibility (MacLean et al., 1998).

Figure 6 shows the links among international co-funders, not including Spanish funders. Although there is a total of 4714 international co-funders, the network presents

Table 9Co-funded publicationsby country (> 19 top 1% pubs)



Fig. 3 Proportion of publications and highly cited publications by country (>19 publications)



Fig. 4 Number of funding countries and proportion of highly cited publications

the 203 nodes that co-fund more than 90 publications. This is one the first visualisations showing an international co-funding network. This network explores a new unit of analysis (funders), also used in Grasano (2017) and Mugabushaka, et al. (2022), to create co-occurrence networks in a similar way to the traditional ones used in co-authorship, co-word or co-citation networks (publications, authors, countries, institutions or keywords).



Fig. 5 Number of funders and proportion of highly cited publications



Fig. 6 International co-funding network

Discussion and conclusions

Surprisingly, only 45.7% of funded projects are mentioned in the FAs of scientific publications. Although this datum is hard to compare because few prior studies have conducted so extensive an analysis, a recent article based on publications about COVID-19 found that only 17% of the papers concerned reported their funding in FAs (Mugabushaka et al., 2022). In addition to technical errors and voluntary and involuntary omissions of funding information, there may be other reasons for the low average, such as the time lag between funding and publishing, funding procedures, or unclear funding agency requirements for acknowledgement of their support (MacLean et al., 1998).

In the analysis of calls, the high percentage of R&D projects is due to the fact that the National Plan implemented a name change a few years ago, eliminating the distinction between fields that existed until 2017. Since then the AEI has not been associating projects with their specialities, and this creates difficulties and obscurity that hamper field-by-field analyses and constrain their scope. The data disaggregated by specialities up to 2017 show that the heaviest investment was made in applied fields, such as Biomedicine, Fundamental Biology, and Agrofood Technologies. The specialities with a higher proportion of projects mentioned in FAs, however, were basic specialities, like Physics, Chemistry, and Mathematics.

It is important to point out that the data restriction other papers have mentioned—referring to the fact that some data sources only include funding acknowledgements since 2015 in Social Sciences and since 2017 in the humanities (Álvarez-Bornstein et al., 2021)—does not affect results in OpenAlex, where papers with FAs have been found dating back to 2011 (referring to projects running on funding granted as long ago as 2008). Therefore, the low percentages of projects mentioned in FAs can be explained by the differing publication habits in these specialities, where publication in scientific journals is not the most frequent channel for disseminating findings.

In the highly cited publications, discipline-related differences in publication and citation are also the most likely explanation. Thus, the lower rate of obsolescence in Mathematics' output could explain the contrast between the high ratios of projects mentioned in FA and funded publications and the low proportion of publications in the top 1%.

The results by regions show that funds are heavily concentrated in a handful of regions. The four regions with the greatest amount of funding account for nearly 75% of all funds invested. This is a clear sign, although not the only sign, that there is a territorial imbalance of human and organisational resources linked to R&D, which helps explain other economic and social phenomena in Spain. On the other hand, the results reveal that high-visibility publications do not depend on the responsible institution's geographic location, but on the quality and prestige of its researchers' output.

The results of the institutional analysis show that, first of all, universities are the institutions with the highest number of funded projects and funds raised (56%). This is not surprising; according to national statistics, 46.8% of researchers work in institutions of higher education, 36.4% work in private companies, and 16.6% work in government agencies (average data for 2008–2020). Secondly, private companies and non-profit organisations receive only 3% of public funds. These data contrast with national socio-economic indicators showing that since 2013 the highest proportion of R&D investment has been private (Fig. 7). This suggests that private R&D in Spain is self-funded. Thirdly, although institutions of higher education are the institutions that receive the most funding, it is the highly specialised centres that present the highest proportion of influential publications.



Fig.7 Rate of Spanish expenditure in R&D by origin (2008–2020). Source: Spanish National Statistics Institute

The results on international co-funding networks show that, as already reported by others (Yegros-Yegros et al., 2013), the preferred partners are the European Union, the United States, and China. They also reveal that both international collaboration and the funding received by the various international partners have a positive influence on the visibility of research output. In other words, the weight of international collaboration plus co-funding is greater in the more-visible and more-influential publications. This increase in representativeness is especially noticeable in the case of publications with co-funding from the European Union, the United States, the United Kingdom, and Germany.

The proportion of highly cited publications increases as the number of funding countries rises. Likewise, the proportion of publications in the top 1%, although lower, also increases as the number of funders grows. Highly cited publications are therefore influenced by both the funders' nationality and the funders' number, confirming previous findings (Gok et al., 2015; Grassano et al., 2017; Zhao, 2010), which could support the idea that the number of co-funders can be regarded as an influence indicator (Boyack et al., 2011). However, these findings differ from those of previous studies showing that the number of funders has a weak influence on visibility (Fortin et al., 2013), or asserting that, due to the weak relationship between visibility and the number of funders, funding bodies do not add anything to the quality of research through their peer-review processes (Rigby, 2011).

Practical implications for research policies

Learning the characteristics, the quantity, the quality, and the use that can be made of funding data in the sources used here is a step forward in FA research. This paper's methodology and results provide an example of the role that this kind of analysis can play to enable researchers, professionals, and funding agencies to evaluate research funding performance effectively.

ACKNOWLEDGEMENTS

We thank the KEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; and the KEK computer group, the National Institute of Informatics, and the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Sciences Laboratory (EMSL) computing group for valuable computing and Science Information NETwork 5 (SINET5) network support. We acknowledge support from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, the Japan Society for the Promotion of Science (JSPS), and the Tau-Lepton Physics Research Center of Nagoya University: the Australian Research Council including Grants No. DP180102629, No. DP170102389, No. DP170102204, No. DP150103061, No. FT130100303: Austrian Science Fund under Grant No. P 26794-N20: the National Natural Science Foundation of China under Contracts No. 11435013, No. 11475187, No. 11521505, No. 11575017, No. 11675166, No. 11705209; Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS), Grant No. QYZDJ-SSW-SLH011; the CAS Center for Excellence in Particle Physics (CCEPP); Fudan University Grants No. JIH5913023, No. IDH5913011/003, No. JIH5913024, No. IDH5913011/002; the Ministry of Education, Youth and Sports of the Czech Republic under Contract No. LTT17020; the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft, the Excellence Cluster Universe, and the VolkswagenStiftung; the Department of Science and Technology of India; the Istituto Nazionale di Fisica Nucleare of Italy; National Research Foundation (NRF) of Korea Grants No. 2014R1A2A2A01005286, No. 2015R1A2A2A01003280, No. 2015H1A2A1033649, No. 2016R1D1A1B01010135, No. 2016K1A3A7A09005603, and No. 2016R1D1A1B02012900; Radiation Science Research Institute, Foreign Large-size Research Facility Application Supporting project and the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information; the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Education and Science of the Russian Federation and the Russian Foundation for Basic Research; the Slovenian Research Agency; Ikerbasque, Basque Foundation for Science, Basque Government (No. IT956-16) and Ministry of Economy and Competitiveness (MINECO) (Juan de la Cierva), Spain; the Swiss National Science Foundation; the Ministry of Education and the Ministry of Science and Technology of Taiwan; and the United States Department of Energy and the National Science Foundation.

Fig. 8 Example of the omission of project identifiers

Experts and analysts from funding agencies especially will find in this methodology a valuable instrument for enhancing the quality and efficacy of their activities, complying with transparency and accountability requirements, facilitating the establishment of objectives that are clearer and easier to measure, and quantifying the scope of the results of funding.

The methodology presented here is a basic tool for evaluating compliance with leading international agencies' self-declared missions and objectives and for locating funding programmes' strong points, but above all it is an excellent tool for finding the weak points that need improvement, strengthening, or change. Otherwise, how can funding agencies hope to learn when the research they fund is innovative, or when their outputs have spread to society, or when the research is considered excellent, or whether they have reached the goal of promoting research and stimulating innovation through interdisciplinary work, to name just some of the objectives agencies themselves claim to seek?

The actors involved must also take action to improve data quality and availability, for results that are more accurate, reliable, and useful. One of the most-sensitive mistakes authors make is to omit precise identification of the organisation and/or project that has funded their research (Fig. 8). As has been indicated, nearly half the publications correctly assigned to the National Plan bore no project identifier.

More and more funding agencies are making it one of their requirements for researchers to include a clear, explicit statement of this information in their research outputs. Few funders, however, are offering advice or instructions on how to do so (Álvarez-Bornstein et al., 2017). Perhaps the simplest, most-effective solution for improving normalisation and reducing ambiguity is to provide an example of how to word this information in the acknowledgements. The AEI itself did not give any express instructions to mention project identifiers until its 2020 call.

Lastly, the creation of a worldwide database of the leading indicators and identifiers of projects funded by public organisations and agencies would be highly useful. This would facilitate matters in the study and evaluation of the tasks involved in funding and scientific publication with open metadata, as well as the production of transparent, reproducible, reusable, comparable, independent results (Eck et al., 2022).

Limitations and further research

This paper does not have the disadvantages associated with sample analysis, because it analyses all the publications with FAs in Crossref and all the projects funded by the Spanish National Plan since 2008. Nonetheless, it is not free of limitations. The Spanish National Plan is the national central administration's instrument for implementing and achieving the objectives of the Spanish Strategy for Science and Technology and for Innovation. The introduction to this paper states some of the main drawbacks that restrict the scope of all results, including the results of this study, such as poor access to quality funder data, problems in detection and availability of the metadata of funder awards in databases, and author errors and omissions in funding reporting.

Moreover, FA-based evaluation is an approach for studying only one facet of the complex work of research. It fails to give insight into other essential aspects of funded scientific activity, such as staff mobility, patents registered, collaboration agreements, contracts signed, training for young researchers, conference organisation, and the acquisition and construction of scientific equipment.

In addition, although the methods, techniques, and results presented in this study are a great help for evaluating funding systems, they cannot replace expert judgement in decision making.

Lastly, this proposal has presented some examples of the kinds of analyses that linked open data make possible, to expand the scope of quantitative FA analysis. Future work testing the usefulness, robustness, and reproducibility of the proposed methods should expand the field of study to include other funders, make a more in-depth analysis of the role of funding agencies in highly cited publications, evaluate the influence of co-authorship and co-funding on publication visibility and productivity, and analyse cofunding networks using additional units of analysis, such as organisations or countries.

Acknowledgements This work was partially supported by Comunidad de Madrid under the Multiannual Agreement with UC3M Grant No. EPUC3M02. The doctoral dissertation of N.R.D. is partially funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Grant No. 88887.687378/2022-00. A previous version of this paper was accepted and presented at the *19th International Conference on Scientometrics and Informetrics* (ISSI2023). Authors would like to thank three anonymous reviewers at ISSI and two anonymous reviewers at Scientometrics for their helpful and valuable comments, corrections and suggestions.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- AEI. (2022). Presentation. [Online]. Madrid: Agencia Estatal de Investigación. https://www.aei.gob.es/en/about-aei/presentation>. [Accessed: 03/01/2024].
- Álvarez-Bornstein, B., & Bordons, M. (2021). Is funding related to higher research impact? Exploring its relationship and the mediating role of collaboration in several disciplines. *Journal of Informetrics*, 15(1), 101102. https://doi.org/10.1016/j.joi.2020.101102
- Álvarez-Bornstein, B., & Montesi, M. (2020). Funding acknowledgements in scientific publications: A literature review. *Research Evaluation*, 29(4), 469–488. https://doi.org/10.1093/reseval/rvaa038
- Álvarez-Bornstein, B., Morillo, F., & Bordons, M. (2017). Funding acknowledgments in the Web of Science: Completeness and accuracy of collected data. *Scientometrics*, 112(3), 1793–1812. https://doi. org/10.1007/s11192-017-2453-4
- ANR. (2022). Missions. [Online]. Paris: Agence National de la Recherche. <<u>https://anr.fr/en/anrs-role-in-research/about-us/missions/></u>. [Accessed: 03/01/2024].
- Boyack KW. (2009). Linking grants to articles: characterization of NIH grant information indexed in Medline. [Online]. Proceedings of ISSI 2009, p. 730–741. https://www.issi-society.org/proce-edings/issi_2009/ISSI2009-proc-vol2_Aug2009_batch1-paper-22.pdf>. [Accessed: 03/01/2024].
- Boyack, K. W., & Borner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the Ameri*can Society for Information Science and Technology, 54(5), 447–461. https://doi.org/10.1002/asi. 10230
- Boyack, K. W., & Jordan, P. (2011). Metrics Associated with NIH Funding: A high-level view. Journal of the American Medical Informatics Association, 18(4), 423–431. https://doi.org/10.1136/amiaj nl-2011-000213
- Butler, L. (2001). Revisiting bibliometric issues using new empirical data. *Research Evaluation*, 10(1), 59–65. https://doi.org/10.3152/147154401781777141
- Clarivate. (2018). InCites. Indicators handbook. [Online]. Philadelphia: Clarivate. https://tinyurl.com/2n37d44e>. [Accessed: 03/01/2024].
- Cronin, B., & Weaver, S. (1995). The praxis of acknowledgment: From bibliometrics to influmetrics. *Revista Española De Documentacion Científica*, 18(2), 172–177. https://doi.org/10.3989/redc.1995.v18.i2.654
- ERC. (2022). ERC at a glance. [Online]. Brussels: European Research Council. https://erc.europa.eu/about-erc/erc-glance>. [Accessed: 03/01/2024].
- Fortin, J. M., & Currie, D. J. (2013). Big science vs. little science: How scientific impact scales with funding. PLoS ONE, 8, e65263. https://doi.org/10.1371/journal.pone.0065263
- Gok, A., Rigby, J., & Shapira, P. (2015). The impact of research funding on scientific outputs: Evidence from six smaller European countries. *Journal of the Association for Information Science and Technol*ogy, 67(3), 715–730. https://doi.org/10.1002/asi.23406
- 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
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). Análise multivariada de dados. Bookman.
- Kramer, B., & de Jonge, H. (2022). The availability and completeness of open funder metadata: Case study for publications funded by the Dutch Research Council. *Quantitative Science Studies*, 3(3), 583–599. https://doi.org/10.1162/qss_a_00210
- Lepori, B., Jongbloed, B., & Hicks, D. (2023). Introduction. Understanding vertical and horizontal complexities in public research funding. In B. Lepori, B. Jongbloed, & D. Hicks (Eds.), *Handbook of public research funding*. Edward Elgar. https://doi.org/10.4337/9781800883086.00005

- MacLean, M., Davies, C., Lewison, G., & Anderson, J. (1998). Evaluating the research activity and impact of funding agencies. *Research Evaluation*, 7(1), 7–16. https://doi.org/10.1093/rev/7.1.7
- Mugabushaka AM. (2020). Linking Publications to funding at project level: a curated dataset of publications reported by FP7 projects. arXiv. DOI: https://doi.org/10.48550/arXiv.2011.07880.
- Mugabushaka AM. (2022). Practically all Scopus-indexed journals are also indexed in open scholarly infrastructures: a contribution to the assessment of data quality in open scholarly data infrastructures. [Pre-print]. Brussels: ERC. 6 p. DOI: https://doi.org/10.5281/zenodo.7307474.
- Mugabushaka, A. M., van Eck, N. J., & Waltman, L. (2022). Funding COVID-19 research: insights from an exploratory analysis using open data infrastructures. *Quantitative Science Studies*, 3(3), 560–582. https://doi.org/10.1162/qss_a_00212
- NSF. (2022). At a glance. [Online]. Alexandria [Virginia]: National Science Foundation. https://www.nsf.gov/about/glance.jsp. [Accessed: 03/01/2024].
- Paul-Hus, A., Desrochers, N., & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167–182. https://doi. org/10.1007/s11192-016-1953-y
- Perianes-Rodríguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, 10(4), 1178–1195. https://doi. org/10.1016/j.joi.2016.10.006
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *International Conference on Science and Technology Indicators*, 26, 7–9.
- Reis, E. (2001). Estatística multivariada aplicada. Sílabo.
- 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. https://doi.org/10.3152/095820211X13164389670392
- Roberge G, Bédard-Vallée A, Rivest M. (2021). Bibliometrics indicators for the science and Engineeering Indicators 2022. Montreal: Science-Metrix. [Online]. < https://www.science-metrix.com/wp-content/ uploads/2021/10/Technical_Documentation_Bibliometrics_SEI_2022_2021-09-14.pdf>. [Accessed: 03/01/2024].
- ROR. (2022). About Research Organization Registry. [Online]. Oakland: California Digital Library. https://ror.org/about/>. [Accessed: 03/01/2024].
- Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. Journal of the Association for Information Science and Technology, 68(3), 790–794. https://doi.org/10.1002/asi. 23713
- Thelwall, M., Simrick, S., Viney, I., & van den Besselaar, P. (2023). What is research funding, how does it influence research, and how is it recorded? Key dimensions of variation. *Scientometrics*, 128, 6085– 6106. https://doi.org/10.1007/s11192-023-04836-w
- van Eck, N. J., & Waltman, L. (2022). Crossref as a source of open bibliographic metadata. Centre for Science and Technology Studies. https://doi.org/10.31222/osf.io/smxe5
- Vieira, S. (2003). Bioestatística: tópicos avançados. Testes não paramétricos, testes diagnósticos, medidas de associação e concordância. Elsevier.
- Yegros-Yegros A, Costas R. (2013). Analysis of the web of science funding acknowledgement information for the design of indicators on external funding attraction. [Online]. *Proceedings of ISSI 2013*, p. 84–95. https://www.issi-society.org/proceedings/issi_2013/ISSI_Proceedings_Volume_I.pdf>. [Accessed: 03/01/2024].
- Zhao, D. (2010). Characteristics and impact of grant-funded research: A case study of the library and information science field. *Scientometrics*, 84(2), 293–306. https://doi.org/10.1007/s11192-010-0191-y

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.