


A Comprehensive Dataset of the Spanish Research Output and Its Associated Social Media and Altmetric Mentions (2016–2020)

Wenceslao Arroyo-Machado , Nicolas Robinson-Garcia  and Daniel Torres-Salinas *

Department of Information and Communication, Faculty of Communication and Documentation, University of Granada, 18071 Granada, Spain; wences@ugr.es (W.A.-M.); elrobin@ugr.es (N.R.-G.)

* Correspondence: torressalinas@ugr.es

Abstract: This paper presents data on research publications authored by scientists affiliated with Spanish institutions between 2016 and 2020, along with their associated social media and altmetric mentions, and on researchers affiliated with Spanish institutions whose work is highly mentioned on social media and non-academic outlets. The first dataset contains 219,988 records and 24 attributes. Each observation represents a scientific publication (article, review or letter) extracted from the Web of Science database. For each record, we provide bibliographic metadata, its subject area and a battery of altmetric indicators extracted from Altmetric.com. The second dataset includes 4209 records and four attributes. Each record corresponds to a researcher. For each record, we include their full name, an author identifier (ORCID), their affiliation and their list of publications connecting to the first dataset.

Dataset: <https://doi.org/10.6084/m9.figshare.19204686>.

Dataset License: CC-BY.

Keywords: altmetrics; publications; Spain; bibliometrics; InfluScience



Citation: Arroyo-Machado, W.; Robinson-Garcia, N.; Torres-Salinas, D. A Comprehensive Dataset of the Spanish Research Output and Its Associated Social Media and Altmetric Mentions (2016–2020). *Data* **2022**, *7*, 59. <https://doi.org/10.3390/data7050059>

Academic Editor: Maarten Marx

Received: 12 March 2022

Accepted: 5 May 2022

Published: 7 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Summary

Altmetrics, alternative metrics based on mentions of scientific publications in social media [1], have been proposed for research evaluation [2]. They still have a long way to go, as there are several limitations attributed to their use [3,4]. However, they offer a different perspective to that offered by citations and can potentially inform on scientific literature consumption beyond academia. The research project “InfluScience—Scientists with social influence: a model to measure knowledge transfer in the digital society” (<https://influscience.eu/>, accessed on 11 March 2022) was launched with the aim to explore the potential of altmetrics, and to study the social influence of Spanish researchers.

As a result of this project, two datasets have been generated in tab-separated values (tsv) format. The first one includes scientific publications authored by scientists affiliated with Spanish institutions between 2016 and 2020 that were retrieved from Web of Science and InCites, thematically classified, and with their altmetric mentions retrieved from Altmetric.com. The most influential Spanish researchers are included in the second one.

Using statistical methods, Spanish scientific activity and the attention received in social media can be studied to identify patterns, trends and distributions within the different metrics, differentiating by scientific areas. These data are of interest to researchers working on scientometrics, altmetrics, science of science or science communication interested on analyzing bibliometric and altmetric production at the macro level.

2. Methods

Data were collected from Web of Science, InCites and Altmetric.com on the 3 March 2021. We first downloaded the publications (articles, editorial material, letters, and proceedings papers) in which an author with Spanish affiliation is listed published between 2016 and 2020 from Web

of Science using the search field Address. This query was limited to the main citation indexes in the Web of Science Core Collection: Science Citation Index Expanded (SCI-Expanded), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), and Emerging Sources Citation Index (ESCI). A total of 434,827 records were downloaded and exported to InCites in order to reclassify records categorized as multidisciplinary in Web of Science. Not all publications were reclassified, and 1171 publications had to be assigned manually to a specific Web of Science subject category.

After reassigning multidisciplinary publications to specific categories, these were classified into the 22 research fields included in the Essential Science Indicators (ESI). We created an equivalence scheme in which each of the 254 subject categories from Web of Science were matched with one ESI field [5]. This classification was conducted following the schema proposed by Tan [6]. Subject categories included in the A&HCI are not integrated into the ESI classification, so we included arts and humanities as an extra research field.

To retrieve altmetric mentions, we used the Digital Object Identifier (DOI) assigned to each publication to query Altmetric.com and obtain all tracked publications and their mentions. From 406,621 records that included a DOI (93.51% of the total), 238,508 were indexed in Altmetric.com (54.85% of the total).

3. Data Description

3.1. Publications Dataset

The first dataset contains scientific publications in which there is at least one author with Spanish affiliation who was mentioned at least once in the Altmetric.com database in the 2016–2020 period. It includes 219,988 records, each being a publication, and 25 variables, including bibliographic information and mentions received. Each publication is assigned to a subject area based on the ESI schema (Essential Science Indicators) provided by Clarivate. Additionally, we include the Arts and Humanities subject area. In total, 18 variables including altmetric indicators are provided: an aggregated score (Influscore) and 17 indicators corresponding to different social and non-academic sources (e.g., Twitter mentions, Facebook mentions, news media). The Influscore is the Altmetric Attention Score (AAS) [7] provided by Altmetric.com on 3 March 2021. These variables are detailed in Table 1. Figure 1 shows the volume of publications retrieved with and without altmetric mentions differentiated by ESI field to reflect the coverage of the dataset provided.

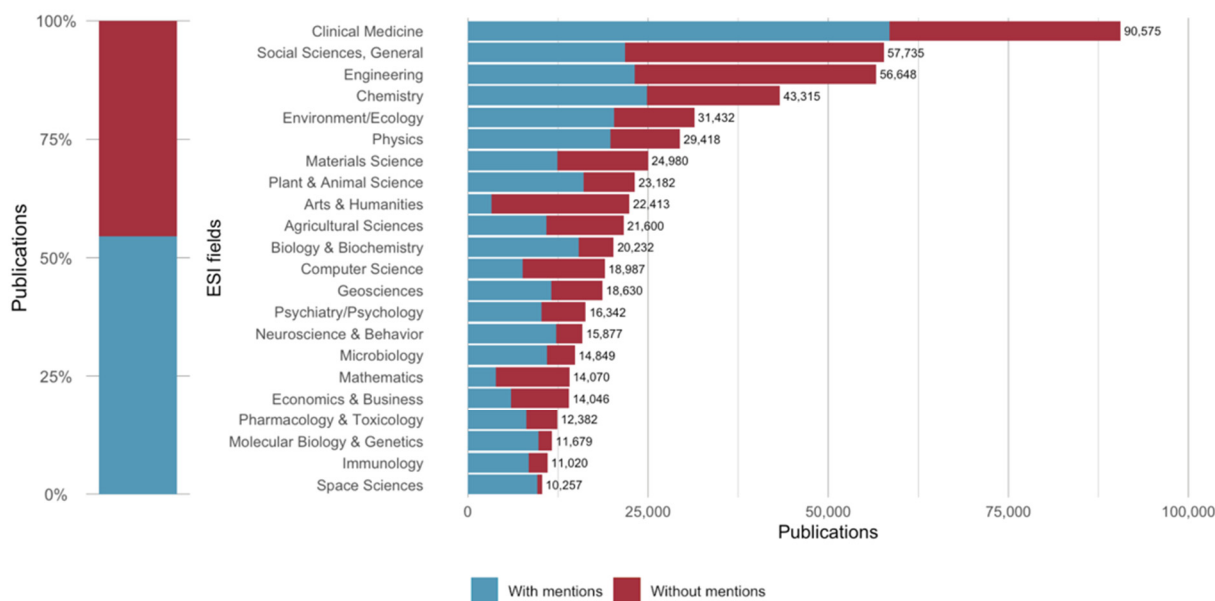


Figure 1. Description of the variables of the publication dataset. Distribution of publications of Spanish researchers (2016–2020) with citations by ESI field.

Table 1. Description of the variables of the publication dataset.

Variable	Type	Description
<i>id</i>	Integer	Unique publication identifier
<i>title</i>	Character	Full title of the publication
<i>year</i>	Integer	Year of publication
<i>type</i>	Character	Document type
<i>journal</i>	Integer	Name of the journal
<i>esi</i> ¹	Character	ESI category of the publication
<i>influcore</i>	Integer	AAS value on 3 March 2021
<i>news</i>	Integer	Number of mentions in news media
<i>blogs</i>	Integer	Number of mentions in blogs
<i>policy</i>	Integer	Number of mentions in policy reports
<i>patent</i>	Integer	Number of mentions in patent
<i>twitter</i>	Integer	Number of mentions in Twitter
<i>post_peer</i>	Integer	Number of mentions in PubPeer and Publons
<i>weibo</i> ²	Integer	Number of mentions in Weibo
<i>facebook</i>	Integer	Number of mentions in Facebook
<i>wikipedia</i>	Integer	Number of mentions in Wikipedia
<i>google</i> ²	Integer	Number of mentions in Google+
<i>linkedin</i> ²	Integer	Number of mentions in LinkedIn
<i>reddit</i>	Integer	Number of mentions in Reddit
<i>pinterest</i> ²	Integer	Number of mentions in Pinterest
<i>f1000</i>	Integer	Number of mentions in F1000
<i>stack_overflow</i>	Integer	Number of mentions in Stack Overflow
<i>youtube</i>	Integer	Number of mentions in YouTube
<i>syllabus</i>	Integer	Number of mentions in Open Syllabus Project

¹ In cases of being more than one category, variables are separated by semicolons. ² Only historical data.

3.2. Authors Dataset

The second dataset includes the top 250 most influential authors at the general level and for each of the ESI fields based on their Influcore. Their information has been reviewed and normalized to produce the Influcore ranking (<https://ranking.influcore.eu/>, accessed on 11 March 2022). For the disambiguation of authors and institutions, we used the algorithm proposed by Caron and van Eck [8]. The dataset is composed of a total of 4209 observations, each being a researcher affiliated to a Spanish institution, and four variables including bio data and linking to the publication dataset. The variables of the second dataset are detailed in Table 2.

Table 2. Description of the variables of the publication dataset.

Variable	Type	Description
<i>name</i>	Character	Full name of the researcher
<i>orcid</i>	Character	ORCID record
<i>organization</i>	Character	Name of the institution of affiliation
<i>publications</i> ¹	Character	List of publication identifiers (id) connecting with the publication dataset

¹ In cases of being in more than one category, variables are separated by semicolons.

Author Contributions: Conceptualization, D.T.-S.; methodology, D.T.-S.; validation, N.R.-G.; data curation, W.A.-M.; writing—original draft preparation, W.A.-M.; writing—review and editing, D.T.-S. and N.R.-G.; visualization, W.A.-M.; supervision, N.R.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This work has funded by the Spanish Ministry of Science and Innovation grant numbers PID2019-109127RB-I00/SRA/10.13039/501100011033 and PID2020-117007RA-I00, and Regional Government of Andalusia Junta de Andalucía grant number A-SEJ-638-UGR20. Wenceslao Arroyo-Machado has an FPU Grant (FPU18/05835) from the Spanish Ministry of Universities. Daniel Torres-Salinas is supported by the Reincorporation Programme for Young Researchers from the Uni-

versity of Granada. Nicolas Robinson-Garcia is funded by a Ramón y Cajal grant from the Spanish Ministry of Science and Innovation (REF: RYC2019-027886-I). Data derived from Altmetric.com were acquired through a research license and are provided under the clauses of such license.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is openly available in figshare at <https://doi.org/10.6084/m9.figshare.19204686>. It does not include any data directly extracted from any Clarivate platform (e.g., Web of Science) as they have only been used as intermediary sources to identify publications authored by Spanish researchers and then to recover their mentions from Altmetric.com. In this sense, the authors' names of the publications have been normalized and clustered and a new paper subject classification has been implemented. With regard to Altmetric.com, a licensing agreement allows us to share this dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Priem, J.; Taraborelli, D.; Groth, P.; Neylon, C. Altmetrics: A Manifesto. Available online: <http://altmetrics.org/manifesto/> (accessed on 11 March 2022).
2. Wouters, P.; Zahedi, Z.; Costas, R. Social Media Metrics for New Research Evaluation. In *Springer Handbook of Science and Technology Indicators*; Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M., Eds.; Springer Handbooks; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 687–713. ISBN 978-3-030-02511-3.
3. Sugimoto, C.R.; Work, S.; Larivière, V.; Haustein, S. Scholarly Use of Social Media and Altmetrics: A Review of the Literature. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 2037–2062. [[CrossRef](#)]
4. Robinson-Garcia, N.; Costas, R.; Isett, K.; Melkers, J.; Hicks, D. The Unbearable Emptiness of Tweeting—About Journal Articles. *PLoS ONE* **2017**, *12*, e0183551. [[CrossRef](#)] [[PubMed](#)]
5. Arroyo-Machado, W.; Torres-Salinas, D. Web of Science Categories (WC, SC, Main Categories) and ESI Disciplines Mapping; 2021. Available online: https://figshare.com/articles/dataset/Web_of_Science_categories_WC_SC_main_categories_and_ESI_disciplines_mapping/14695176/2 (accessed on 11 March 2022). [[CrossRef](#)]
6. Tan, F. Mapping of Subject Category (SC) Field—ESI Discipline Category; 2020. Available online: https://figshare.com/articles/dataset/Mapping_of_Subject_category_SC_field_-_ESI_discipline_category/13269737 (accessed on 11 March 2022). [[CrossRef](#)]
7. Elmore, S.A. The Altmetric Attention Score: What Does It Mean and Why Should I Care? *Toxicol. Pathol.* **2018**, *46*, 252–255. [[CrossRef](#)]
8. Caron, E.; van Eck, N.-J. Large Scale Author Name Disambiguation Using Rule-Based Scoring and Clustering: International Conference on Science and Technology Indicators. In *Proceedings of the Science and Technology Indicators Conference, Leiden, The Netherlands, 3–5 September 2014*; pp. 79–86.