Use of Research Organizations Registry (ROR) identifiers in author academic profiles: the case of Google Scholar Profiles

ENRIQUE ORDUÑA-MALEA
Universitat Politècnica de València
enorma@upv.es
https://orcid.org/0000-0002-1989-8477

NÚRIA BAUTISTA-PUIG
Universidad Complutense de Madrid
nuriabau@ucm.es
https://orcid.org/0000-0003-2404-0683

Uso de identificadores del Registro de Organizaciones de Investigación (ROR) en perfiles académicos de autor: el caso de Google Scholar Profiles

ABSTRACT

Research organizations' persistent identifiers allow for reducing affiliation ambiguities, enable accurate institutional analyses and favor the design of modern online scholarly databases suited for research discovery and research evaluation. However, few studies have attempted to quantify their degree of use. The purpose of this work is precisely to determine the use of Research Organizations Registry (ROR) IDs in author academic profiles, specifically in Google Scholar Profiles (GSP). To do this, all the Google Scholar profiles including the term ROR in any of the public descriptive fields were collected and analyzed. The results evidence a low use of ROR IDs (1,033 profiles), mainly from a few institutions (e.g. *Pontificia Universidad Javeriana* in Colombia, and *Escuela Superior Politécnica del Litoral* in Ecuador hold 55.7% of all profiles), from low citation-based impact authors (45.1% of profiles attain less than 100 citations each), belonging mainly to Social Sciences (26.3%), Engineering fields (25.3%), and Natural Sciences (22.2%). Although Google Scholar does not facilitate the inclusion of identifiers, it seems that the world's leading research institutions are not recommending their researchers include these identifiers in their profiles yet.

RESUMEN

Los identificadores persistentes de las organizaciones de investigación permiten reducir las ambigüedades de afiliaciones, realizar análisis institucionales precisos y favorecen el diseño de modernas bases de datos online *académicas adecuadas para el descubrimiento y la evaluación la investigación. Sin embargo, pocos estudios han intentado cuantificar su grado de utilización. El propósito de este trabajo es precisamente determinar el uso de los ID del Registro de Organizaciones de Investigación (ROR) en los perfiles académicos de los autores, específicamente en los perfiles de Google Scholar (GSP). Para ello, se recopilaron y analizaron todos los perfiles de Google Scholar que incluían el término ROR en alguno de los campos descriptivos públicos. Los resultados evidencian un bajo uso de identificadores ROR (1.033 perfiles), principalmente de pocas instituciones (por ejemplo, la* Pontificia Universidad Javeriana *en Colombia y la* Escuela Superior Politécnica del Litoral *en Ecuador tienen el 55,7% de todos los perfiles), de autores de bajo impacto basado en citas (el 45,1% de los perfiles alcanzan menos de 100 citas cada uno), pertenecientes principalmente a los campos de Ciencias Sociales (26,3%), Ingenierías (25,3%) y Ciencias Naturales (22,2%). Aunque Google Scholar no facilita la incorporación de identificadores, parece que las principales instituciones de investigación del mundo aún no recomiendan a su personal investigador la inclusión de estos identificadores en sus perfiles.*

KEYWORDS

Google Scholar; Research organizations registry; Author profiles; Academic search engines; Persistent identifiers; Scientometrics; Meta-research.

PALABRAS CLAVE

Google Académico; Registro de organismos de investigación; Perfiles de autor; Motores de búsqueda académicos; Identificadores persistentes; Cienciometría; Meta-investigación.

Ús d'identificadors del registre d'organitzacions de recerca (ROR) als perfils d'autor: el cas de *Google Scholar Profiles*

RESUM

Els identificadors persistents de les organitzacions de recerca permeten reduir les ambigüitats de les afiliacions, la realització d'anàlisis institucionals precisos i afavoreixen el disseny de bases de dades *online* acadèmiques modernes adequades per al descobriment i l'avaluació de la recerca. No obstant això, pocs estudis han intentat quantificar el seu grau d'utilització. L'objectiu d'aquest treball és precisament determinar l'ús dels identificadors del registre d'organitzacions de recerca (ROR) als perfils acadèmics d'autor, concretament als perfils de Google Scholar (GSP). Per fer-ho, es van recollir i analitzar tots els perfils de Google Scholar que inclouen el terme ROR en qualsevol dels camps descriptius públics. Els resultats evidencien un baix ús del ROR ID (1.033 perfils), per part de poques institucions (per exemple, la *Pontificia Universidad Javeriana* de Colòmbia i l'*Escola Superior Politècnica del Litoral* de l'Equador tenen el 55,7% de tots els perfils), d'autors de baix impacte basat en citacions (el 45,1% dels perfils tenen menys de 100 citacions cadascun), pertanyent principalment a les àrees de Ciències Socials (26,3%), Enginyeria (25,3%) i Ciències Naturals (22,2%). Tot i que Google Scholar no facilita la inclusió d'identificadors, sembla que les institucions de recerca líders del món encara no recomanen als seus investigadors que incloguin aquests identificadors als seus perfils.

PARAULES CLAU

Google Scholar; Registre d'entitats de recerca; Perfils d'autor; Motors de cerca acadèmics; Identificadors persistents; Cienciometria; Meta-recerca.

1. Introduction

ROR (Research Organization Registry)[1] is a community-led project launched in 2019 that attempts to create standard identifiers for research organizations (Meadows, 2019).

Its origins date back to series of meetings held between 2016 and 2018 by a large (17) number of organizations, in which the need for a solution providing "resolvable, persistent and unique identifiers for organizations involved in research that can be used to describe researcher affiliations" was raised (Demeranville et al., 2016). Both a top-down (a working group defining governance recommendations and product principles)[2] and a bottom-up (a Request for Information)[3] processes were initiated to define a proposal. Finally, the California Digital Library, Crossref, DataCite, and Digital Science shaped a new steering group to implement the proposal, with a donation of seed data from Digital Science's GRID proprietary database.

Since the public retirement of GRID in September 2021, from whose database was ROR initially fed, ROR started to be maintained independently as a leading "open, stakeholder-governed infrastructure for research organization identifiers and their associated metadata",[4] covering more than 102,815 registries as of 28 August 2022.

As other research organizations identifiers (e.g., Crossref Funder ID, GRID, ISNI, Ringgold Org ID, Wikidata), ROR helps reducing affiliation ambiguities while allow accounting and characterizing research organizations worldwide. Moreover, the fact that ROR IDs are also embedded in URLs[5] eases online navigation between research objects (e.g., journal articles, bibliographic records) and organizations' metadata, and enables webometric studies and link-based analyses.

Future full open integration between publications (e.g., DOIs and Handles), authors (e.g., ORCIDs) and organizations identifiers (e.g., RORs) would also allow the design of modern hybrid scholarly databases, academic platforms and value-added online services which will favor both research discoverability and research evaluation exercises.

Studies related to the degree of use of ROR IDs are needed to understand how this standard is being incorporated in the ecosystem of science. Specifically, discovering the main objects incorporating ROR IDs (e.g., publications, authors, publishers, and institutions) and testing new ROR-based indicators would facilitate a better understanding of the (economic, technical, and academic) benefits of this new standard, and would help calibrating the feasibility of new academic-related products design.

Orduña-Malea & Bautista-Puig (2022) carried out the first large-scale study of ROR IDs by analyzing 147,154 links targeted to 102,559 ROR identifiers, finding that the percentage of ROR

IDs linked is still limited (only 51.6% of ROR identifiers have been linked at least once). The link-based ROR IDs were mainly placed as an affiliation field on the publications' bibliographic records (51.4% of links) created by repositories or scholarly databases, while few publications included link-based ROR IDs within the publication itself (as ORCIDs or DOIs already do), mainly in the acknowledgements section.

However, link-based analyses of ROR IDs show a sort of methodological limitations. First, the ROR ID should be embedded in a URL, which prevents the identification of raw ROR IDs in texts. Second, link analysis is constrained by the features of the bots used to collect the linking webpages. Commercial link analysis tools leave behind webpages with specific bot exclusion policies or web accessibility problems as well as social media platforms (e.g., Tweets, Facebook posts, etc.), which are generally closed to external crawlers when it comes to index internal pages (i.e., each profile). This limitation includes links and mentions from author profiles such as ResearchGate or Google Scholar Profiles (also referred to as Google Scholar Citations).

This study aims to fill this gap by analyzing the use of ROR IDs in author academic profiles via mention-related identification, instead of link-related. To do this, the following research questions are set:

- **RQ1.** How many authors do include ROR IDs in their author academic profiles?

- **RQ2.** How are authors including their ROR IDs in their academic profiles?

- **RQ3.** Do highly cited authors include ROR IDs in their academic profiles?

- **RQ4.** Which research organizations the authors who include ROR IDs in their academic profiles belong to?

- **RQ5.** Which research fields the authors who include ROR IDs in their academic profiles belong to?

In order to response properly to the research questions set above, Google Scholar Profiles will be used as an author profile case study.

2. Research background

Google Scholar Profiles (GSP) is a free author academic profile service created by Google in 2011 (Jacsó, 2012). It allows users to collect their publications among the bibliographic data already indexed in Google Scholar, to edit these records (e.g., merging different versions, fixing bibliographic data errors) and to add basic personal information (name, picture, affiliation, topics of interest, coauthors). The profile automatically computes basic author-level metrics (h-index, i10, citations). A complete description of GSP is available in Delgado

López-Cózar, Orduña-Malea & Martín-Martín (2019).

GSP has been extensively analyzed by the literature since its inception. Some studies have focused on its features, such as author metadata (Ortega, 2015a; Orduña-Malea et al., 2017), bibliometric indicators (Teixeira da Silva, 2018; 2021), citation data accuracy (Doğan, Şencan & Tonta, 2016) or even citation manipulation effects (Delgado López-Cózar, Robinson-García & Torres-Salinas, 2014). Other studies have analyzed the coverage of GSP, either global (Ortega & Aguillo, 2014; Ortega, 2015b) or centered on the presence of specific disciplines (Kim & Grofman, 2020), institutions (Mikki et al., 2015; Ortega, 2015c) or research groups (Thoma & Chan, 2019). Finally, other studies have directly used GSP as a source for bibliometric analyses to carry out co-keyword (Ortega & Aguillo, 2012), institutional collaboration (Ortega & Aguillo, 2013) and disciplinary (Martín-Martín, Orduna-Malea & Delgado López-Cózar, 2018) analyses.

Due to its huge coverage and refreshment velocity (the same as Google Scholar), GSP shows advantages for meta-research and Science studies. However, the interactive features offered by this author academic profile are basic. Since 2011 a few updates have been carried out. At the private level, GSP has introduced over the years automatic recommendations profiles, improved the follow colleagues' profiles button (October 2017)[6] and enhanced options for articles' recommendation (e.g., save articles) (February 2021)[7], or automatic suggestions to fix bibliographic errors. At the public level, GSP updated the visual interface in August 2014,[8] included citation histograms over the years (September 2014),[9] disclosed information about public access of publications and allowed a procedure to use Google Drive to upload documents (March 2021),[10] and recently, the possibility of including alternative author names (August 2022).

Among the public-level updates, GSP introduced in August 2015 an affiliation link. While this feature allowed facilitating the identification of authors' affiliation by enabling a hyperlink and creating institutional lists of authors, several errors and shortcomings were pointed out (Orduña-Malea et al., 2017). Precisely, the inclusion of ROR IDs in the GSP profiles would enhance the affiliation information offered, disambiguate institutions, and fix errors.

At present, that possibility is not integrated in the profile, but authors can include the ROR ID within the description field text box. Precisely, this study is intended to shed light on how and to what extend ROR IDs are being used in GSP.

3. Method

To accomplish with the objectives of this study, the GSP search feature was directly used to collect all author profiles including the 'ROR' text chain in their profiles. All the profiles
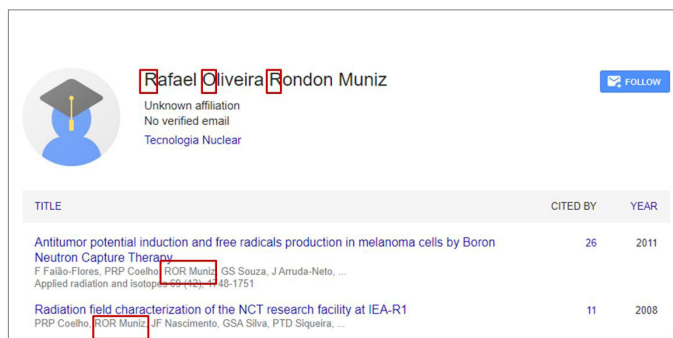
Figure 1. Example of Google Scholar Profile collected via 'ROR' plain search.

were extracted by 19 August 2021 via web scraping and exported to a spreadsheet to further analysis. For each profile, the name (and position description, when available), verified email domain, description, areas of interest (i.e., keywords added by the author) and total number of citations received were gathered.

GSP search does not allow to search in specific fields of each profile (e.g., name, description, verified address), except for author keywords (e.g., label=example). Therefore, a data cleansing process was carried out to filter out false positives.

Only 16 profiles were found to be false positives, a low value. While on some occasions the 'ROR' appeared in the author's name (e.g., Shivani Ror), in other occasions we detected that the name initials do correspond to the chain text (e.g., Rafael Olivera Rondón Muniz). In these cases, Google Scholar wrongly creates a false author name (i.e., ROR Muniz), which contains the text searched. Curiously enough, 'ROR Muniz' (this same happens for other similar cases) does not appear in the author information, but in the publications' bibliographic records included in the profile (Figure 1).

Once the dataset was cleansed, the genre was manually determined for each author profile by using the Genderize app[11]. This tool allows inferring the gender (female/male/

unknown) by using the author's first name (and optionally, indicating the country). Unfortunately, as the information provided by GSP is not accurate enough, only a binary categorization (female/male) was adopted through the author's name. In case of ambiguity, the author picture, when available, was checked (e.g., F. Xavier).

Areas of interest (i.e., keywords) included by each author were subsequently used to determine the authors' disciplines. To do this, the following predefined classification was adopted: *Arts and Human sciences* (e.g. Music, Sculpture, Painting, language studies, History, Philosophy), *Formal sciences* (i.e. Mathematics), *Natural sciences* (e.g. Physics, Chemistry, and Biology), *Applied sciences I* (i.e. all the engineering), *Applied Sciences II* (i.e. Health and Medicine), *Social Sciences* (e.g. Economy, Business, Sociology, Psychology, Law, Tourism, Library and Information Sciences, Communication). When no keywords were included in the profile, the Department information, description field and the publications included were used to determine the discipline.

Finally, each profile was manually examined to identify the ROR ID included, determining the procedure used by authors to embed the identifier, establishing the following categories:

- **Raw ROR ID**: the ID is included without URL format (e.g., ROR 02f40zc51)

- **Domain name**: the ID is included as a domain name (e.g., ror.org/02p0gd045)

- **Full link**: the ID is embedded in a URL (e.g., https://ror.org/012a91z28).

- **Incomplete**: Although the ROR acronym has been added to the profile, the ID is not included (e.g., grid.41312.35 / ROR).
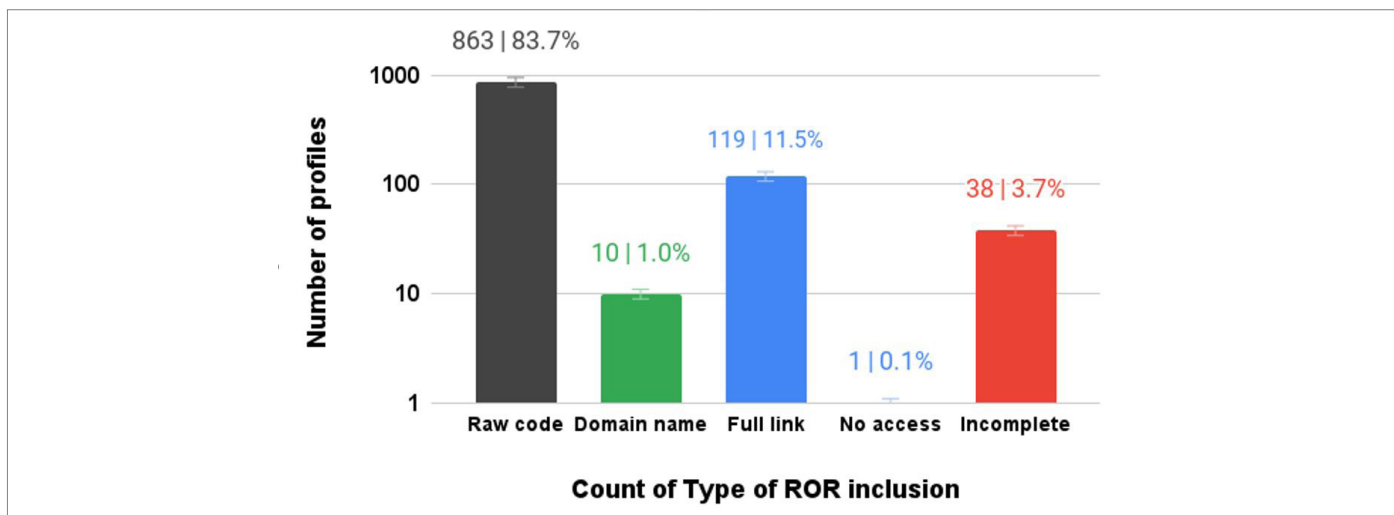

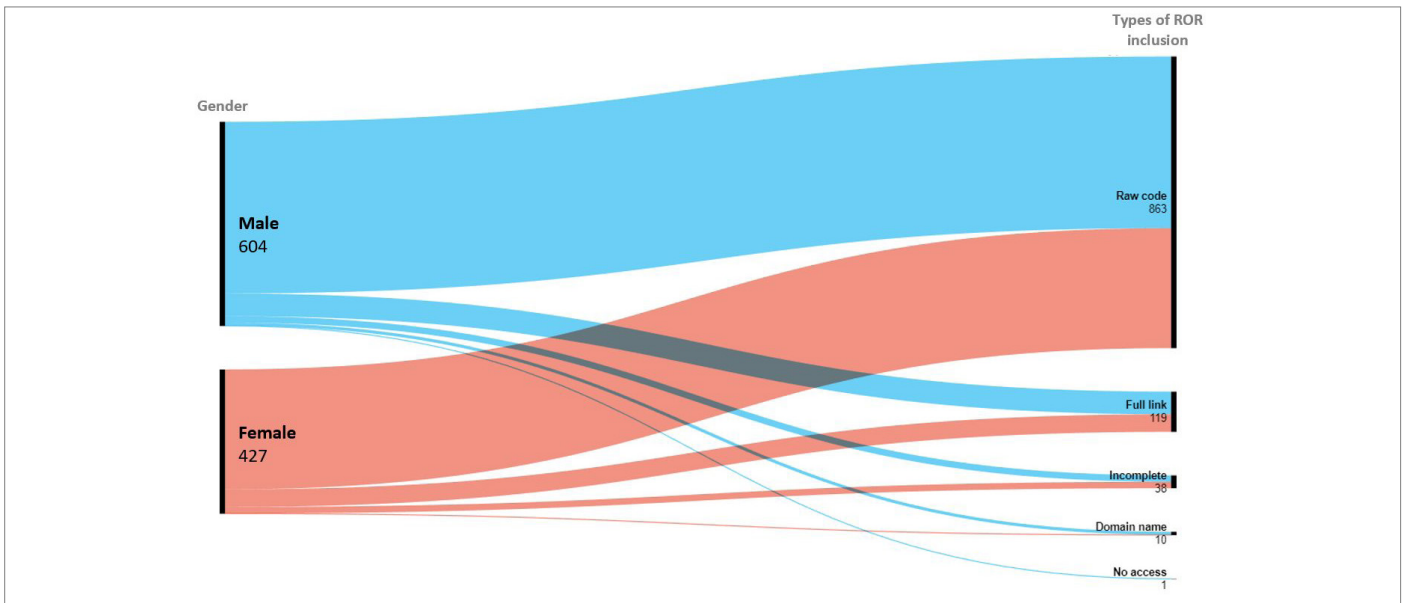**Figure 2.** Count of types of ROR inclusion in Google Scholar Profiles.

Figure 3. Alluvial Diagram of types of ROR inclusion by gender. Note: powered with RAWGraphs, https://www.rawgraphs.io

4. Results

4.1. Authors

1,033 profiles in GSP include a ROR ID. One profile is invented (Nombre Apellido1-Apellido2) and another one is institutional (FCNM ESPOL). Considering the remaining 1,031 author profiles, 58.6% are male authors and 41.4% are female authors.

Most authors (83.7%) have included the ROR ID as a raw code in the description field, while the full link (11.5%) and domain name (1.0%) are minority options. It stands out the remarkable number of profiles (38) omitting the ROR ID, although the acronym has been included (Figure 2). Finally, one profile was not public at the time of the analysis.

From a gender perspective, almost 6 out of 10 men include the ROR ID as raw code in the description field whereas this percentage is lower for women (4 out of 10). Similarly, the proportion of scholars including the ROR ID as a full link is lower for woman (12.61 points lower than men). The incomplete option, in which the ID is mentioned but not included, presents the same value in both groups (19 authors/each) (Figure 3).

The citations average value of the 1,031 author profiles is high (1,024.96 citations per profile). However, the citation distribution is quite skewed. While a significant percentage of authors (19.7%) have attained at least 1,000 citations, 45.1% of authors exhibiting less than 100 citations, including 56 profiles with no citations (Figure 4).
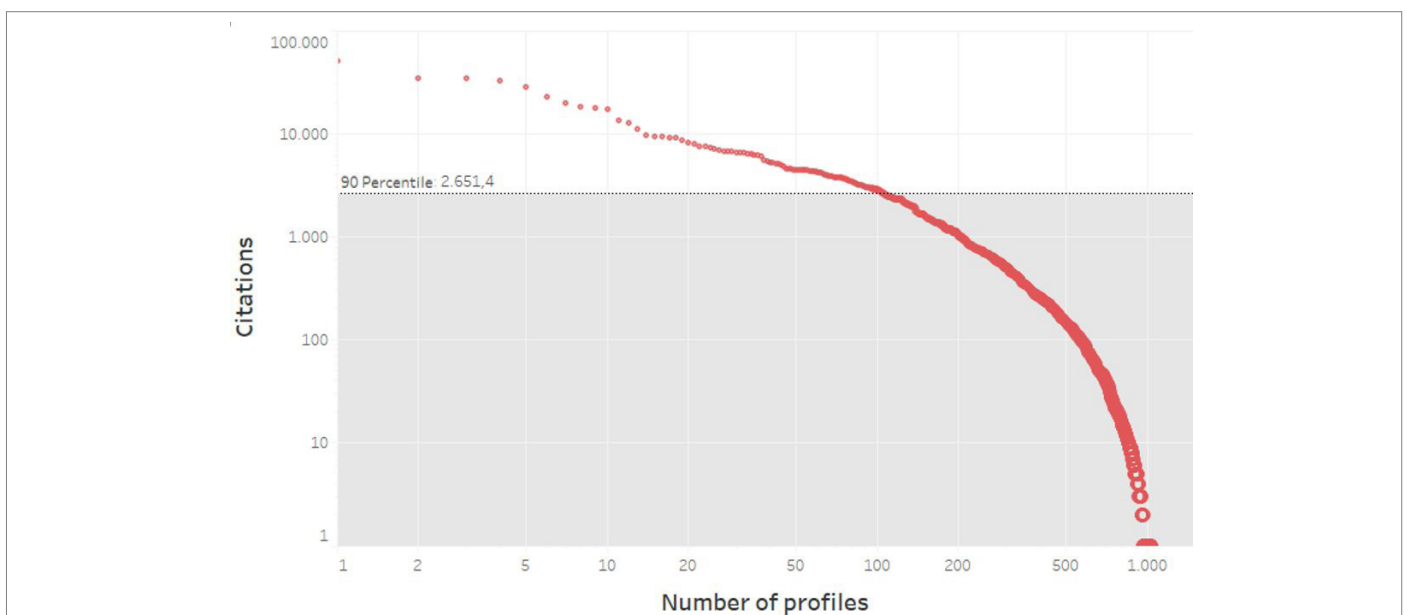

Figure 4. Distribution of citations for Google Scholar Profiles including ROR IDs

Research organization (Domain name)	Country	Number of Profiles	%	Citations (Mean)	Citations (Median)	Citations (Max)
javeriana.edu.co	Colombia	400	39.3	326	96.5	18,399
espol.edu.ec	Ecuador	167	16.4	125	25	1,459
csic.es	Spain	37	3.6	3,041	2,288	19,587
mondragon.edu	Spain	31	3.0	293	49	4,504
um.es	Spain	20	2.0	1,477	156	9,270
uclm.es	Spain	19	1.9	1,619	1,513	5,194
utp.ac.pa	Panama	18	1.8	51	15.5	287
usal.es	Spain	15	1.5	5,200	771	50,398
uniovi.es	Spain	15	1.5	3,424	2,866	12,599
alzahra.ac.ir	Iran	13	1.3	834	181	4,332
ucm.es	Spain	12	1.2	2,070	1,123	6,802
urjc.es	Spain	10	1.0	1,461	870.5	6,967

Table 1. Number of profiles according to the authors' research organization.

4.2. Research organizations

Most of the profiles included institutional affiliation via the email verified domain (1017; 98.6%). A total of 144 research organizations have been identified being The *Pontificia Universidad Javeriana* (javeriana.edu.co) in Colombia the institution with most public profiles in GSP including a ROR ID (400 profiles), followed by the *Escuela Superior Politécnica del Litoral-ESPOL* (espol.edu.ec) in Ecuador with 167 profiles. However, the number of citations attained by the profiles of these institutions is limited (Table 1). The third research institution is the *Consejo Superior de Investigaciones Científicas* (csic.es) in Spain, with a fewer number of profiles (37) but higher median of citations received (2,288). To reflect the prestige of profiles by institution, the maximum citation value is also provided.

The results offered by the Table 1 evidence the prominence of Spanish-speaking research organizations, especially from Spain (8 Spanish institutions among the top 12 institutions most mentioned via ROR IDs from GSP). The unique non-Spanish-speaking institution among the top ten corresponds to the *Alzahra University*, from Iran.

A country-level analysis has revealed the presence of 25 different countries, showing Spain (78 institutions), Colombia (9) and Mexico (7) the countries with most institutions covered (Table 2). These results show a significant presence of Latin-American countries, and an unexpected absence of Anglo-Saxon countries. United States includes two institutions (*Binghamton University* and *Waters Corporation*), while United Kingdom and Australia do not include any profile. European institutions are also absent (Belgium includes two institutions, Switzerland, and Germany one institution each). Ukraine is the first non-Spanish-speaking country (four

institutions). This concentration of profiles including ROR IDs from Spanish-speaking countries could be the result of specific actions carried out to enrich academic profiles, which are still absent in other countries.

Otherwise, the penetration of institutions in each country shows different patterns. For example, while Mexico includes 7 different institutions, it only generates 9 profiles, which means an anecdotic presence. In contrast, Ecuador (via *ESPOL*) or Colombia (via *Pontificia Universidad Javeriana*) reveals the existence of institutions which are creating advanced profiles possibly through research policies' strategies.

Country	Number of research organizations	%	Number of profiles	%
Spain	78	54.2	321	31.5
Colombia	9	6.3	410	40.2
Mexico	7	4.9	9	0.9
Ecuador	6	4.2	173	17.0
Chile	5	3.5	5	0.5
Ukraine	4	2.8	14	1.4
Peru	4	2.8	6	0.6
Panama	4	2.8	24	2.4
Indonesia	3	2.1	7	0.7
Venezuela	3	2.1	3	0.3
United States	2	1.4	2	0.2
Belgium	2	1.4	2	0.2
Norway	2	1.4	2	0.2
Iran	2	1.4	20	2.0
Brazil	2	1.4	4	0.4

Table 2. Number of profiles according to the authors' research organization country (>2)

Out of the 144 research organizations identified, the type of the institution has been identified for 140 institutions (Table 3). The presence of universities (75.7% of all institutions) stands out, followed by far by research institutes (9) and government bodies (6). 52 Spanish universities have at least one profile including a ROR ID.

Research organizations type	Number of research organizations	%
University	106	75.7
Research institute	9	6.4
Government	6	4.3
Hospital	3	2.1
Public Research Organization	3	2.1
Company	2	1.4
Personal website	2	1.4
Research center	2	1.4
Technologic center	2	1.4
Education Institution	1	0.7
Federation	1	0.7
National Academy	1	0.7
Organization	1	0.7
Research Foundation	1	0.7
TOTAL	140	100

Table 3. Type of research organizations

4.3. Research fields

Social Sciences (26.3%), Applied Sciences I-Engineering (25.3%) and Natural Sciences (22.2%) constitute the research disciplines most represented by the authors who have included a ROR IR to their Google Scholar Profile. While the low presence of Mathematics (1.7%) or Human Sciences (11.7%) was expected, the low percentage of Health and Medicine related profiles (12.7%) is remarkable (Figure 5).

Natural Sciences show the highest median value for the number of citations received (500), followed at some distance by the Applied Sciences II–Health and Medicine (140). Otherwise, both Applied Sciences I-Engineering (96) and Social Sciences (108) exhibit lower median citations values, as expected due to the general citation patterns in each of the disciplines.

5. Discussion

This study has analyzed the presence of ROR IDs in Google Scholar Profiles, identifying the quantity of public profiles including this identifier and the mode used to be included. In addition, some characteristics of the authors including ROR IDs (genre, citation-based impact, organization type, country, research field) have been pointed out.

Considering the GSP population, currently estimated around 4 million profiles,[12] the inclusion of ROR IDs in the Google Scholar profiles is considered low. Moreover, these profiles are heavily concentrated in two institutions. A plausible explanation is that the inclusion obeys to specific guidelines provided by *ad hoc* courses on research promotion (e.g., guides from the *Pontificia Universidad Javeriana*)[13] or specific research policies. The fact that 45.1% of authors exhibit less than 100 citations (early-career researchers) and 85.8% of the 1,031 profiles including a ROR ID also include the author's ORCID ID either in the name, description, or website fields (a specific and advanced self-promotion action) reinforces the hypothesis that authors were trained to create a digital research entity, including the ROR identifier.

An unbalanced gender percentage has been found. A lower presence of women in GSP was already pointed out by Tsou et al. (2016). This might indicate that men scholars are more interested in maintaining their profiles than women. These results reinforce the need to carry out gender studies in research promotion activities by universities to better understand the different use of author profiles from a gender
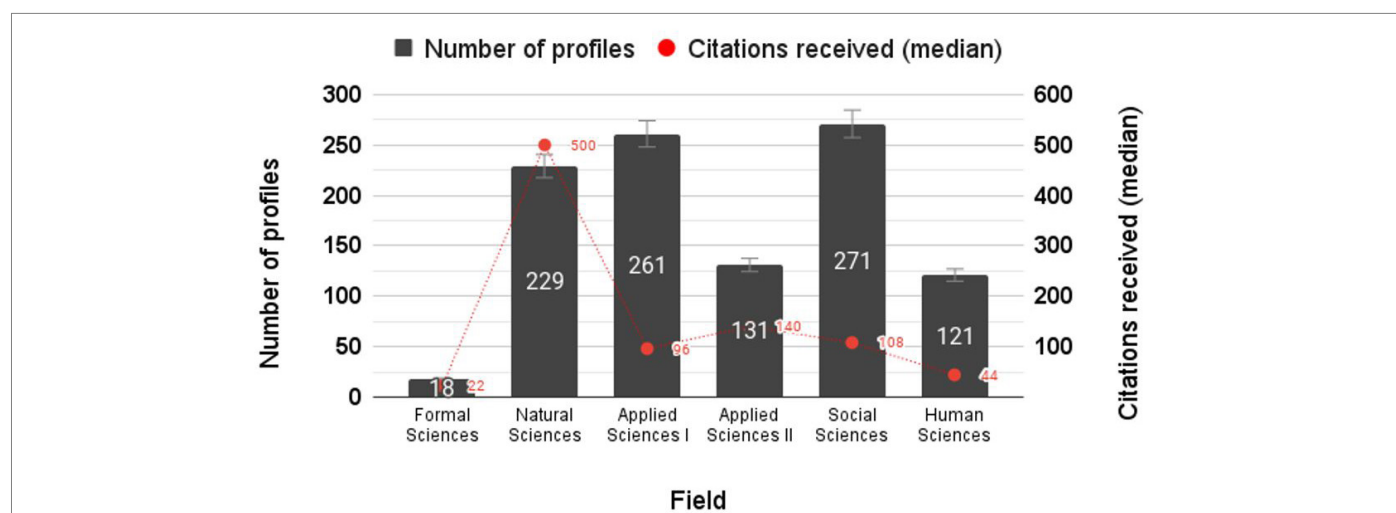

Figure 5. Distribution of Google Scholar Profiles including ROR IDs according to research field and median of citations received.

perspective. In this case, and considering the lower presence of women, similar patterns of ROR ID inclusion have been found. However, the inclusion of ROR IDs as a raw code and a full link (the most frequently mode used in the sample) is significantly higher for men.

The absence of North American, Canadian, North European, Australian, or Asiatic institutions is quite remarkable. Spain, Colombia, and Ecuador provide 88.7% of all the profiles. These results contrast with previous findings pointing out a bias towards Anglo-Saxon countries when measuring webpages linking to URL-based ROR IDs (Orduña-Malea & Bautista-Puig, 2022). The geographic origin of links to ROR IDs is unrelated to the geographic presence of authors including ROR IDs in their Google Scholar Profiles. Arguably, the inclusion of ROR IDs in Google Scholar Profiles is not being promoted in those countries so far. However, the scholarly databases in those places have started to implement ROR IDs in their bibliographic records.

The fact that Google Scholar Profiles do not provide specific fields for identifiers, neither ORCIDs nor RORs, might explain the low penetration level, as authors should include the identifier in other descriptive fields voluntarily. Nonetheless, this is counterproductive as authors introduce the identifiers in fields not designed for that purpose, limiting their properties and utility. We also observe this effect with ORCIDs, which are mainly included in the name field. This is not the best option because the ORCID ID is embedded as a plain text chain as part of the author's name, which is inaccurate and might jeopardize the location of the author as Google does not understand the ORCID ID as a code, but as part of the authors' name. The use of the website field constitutes a better option, as an active link is provided, and the reader can click and access to the author's ORCID profile.

Finally, the high presence of Social Science researchers is also quite remarkable, as the distribution of fields observed in the Figure 4 does not correspond to the publication patterns of scientific results or the size of research communities. Again, the effect of specific training from and to this community might explain the results obtained.

Limitations

The outcomes obtained in this study have come through several research limitations that should be pointed out to discuss and contextualize the results properly.

First, the search method is limited to find the chain 'ROR' in the public descriptive fields of each profile. This strategy prevents the discovery of ROR IDs embedded directly without the ROR acronym. In any case, the authors believe that the strategy used collected most of the existing profiles, being non-significant the number of the omitted profiles.

Second, data relies on the public information provided by the authors in their profiles. However, this information can be incomplete, erroneous or even fake. For example, the total number of citations can be misleading if the profile is not curated (e.g., including publications not authored by the author). While profiles have not been examined to check their accurateness, results have been displayed via aggregated indicators (mean, median), which minimizes the effects on the results.

Third, affiliation data has been collected from the verified email provided by each author. The main advantage of this procedure is that domain names are unique, and this allows disambiguating institutions. However, this limits to one institution per author, arguably the main institution. In addition, public profiles can be published without a verified email. In those cases, no institutional data was collected. However, the number of emails with no verified email in the sample was non-significant (13; 1.7%).

Fourth, research fields were identified through the areas of interest, that is, up to five keywords provided by the authors. Notwithstanding, authors can include typos, inaccurate or misleading terms, or terms unrelated with their publications but their interests. In some cases, authors do not include keywords (22 profiles in the sample do not show areas of interest). In other cases, the department/school mentioned by the author, if any, was misaligned with the keywords included, although research and teaching are not necessarily related in higher education institutions. Additionally, a very generic thematic classification was employed as a pragmatic solution (e.g., the line between engineering and natural sciences is sometimes very thin). For all these reasons, the disciplinary classification should be considered assuming a tolerable margin of error.

Finally, the low use of ROR IDs found in this study is consistent with the results previously shown by Orduña-Malea & Bautista-Puig (2022), who found that still few scholarly databases embedded links to ROR IDs. The authors estimate that data quality differences between ROR and GRID records could limit ROR implementation, and consequently, its popularity among scholars and scientists.

6. Conclusions

This study has revealed a low degree of inclusion of ROR IDs in the Google Scholar Profiles. The use of ROR IDs is limited to few institutions probably because of specific research advisory or training activities. The authors of the most important universities and research institutions worldwide are not including ROR IDs yet in their profiles.

The availability of specific descriptive fields for identifiers (mainly ORCID and ROR) is deemed recommended to the

Google Scholar team to encourage the use of these standards, which might increase the users' browsing experience and the elaboration of more accurate research studies.

Future studies should analyze the degree of use of ROR IDs by monitoring their inclusion in other author profiles and social media platforms, and its web connectivity with ORCIDs and DOIs, which might lead meta-researchers to a new generation of web-based Science studies.

Endnotes

1. https://ror.org
2. https://doi.org/10.23640/07243.12827315.v1
3. https://doi.org/10.23640/07243.5458162.v1
4. https://ror.org/about/
5. For example, the URL https://ror.org/02p0gd045, contains the ROR ID: 02p0gd045, which corresponds to the *Universidad Complutense de Madrid*.
6. https://scholar.googleblog.com/2017/10/follow-related-research-for-key-authors.html
7. https://scholar.googleblog.com/2021/02/scholar-recommendations-reloaded.html
8. https://scholar.googleblog.com/2014/08/fresh-look-of-scholar-profiles.html
9. https://scholar.googleblog.com/2014/09/a-rapid-round-of-ui-changes.html
10. https://scholar.googleblog.com/2021/03/track-and-manage-your-public-access.html
11. https://genderize.io/
12. This data constitutes an estimation considering the 1.2 million profiles managed by the Ranking Web of Researchers (https://webometrics.info/en/node/194), a project carried out by the Cybermetrics Lab (CSIC, Spain). This data has been also discussed in personal conversation with Isidro Aguillo, Head of the Cybermetrics Lab.
13. 'How to create a profile in Google Scholar' guide by Pontificia Universidad Javeriana, https://javeriana.libguides.com/ld.php?content_id=61373607

Acknowledgements

References

Delgado López-Cózar, E., Orduña-Malea, E. & Martín-Martín, A. (2019). Google Scholar as a Data Source for Research Assessment. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 95-127). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_4

Delgado López-Cózar, E., Robinson-García, N. & Torres-Salinas, D. (2014). The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, *65*(3), 446-454. https://doi.org/10.1002/asi.23056

Demeranville, T., Brown, J., Fenner, M., Cruse, P., Haak, L., Paglione, L., Bilder, G., Lin, J. & Pentz (Eds). (2016). *Organisation Identifiers -*

Minimum viable product requirements. https://dx.doi.org/10.6084/m9.figshare.3479141.v1

Doğan, G., Şencan, İ. & Tonta, Y. (2016). Does Dirty Data Affect Google Scholar Citations? *Proceedings of the Association for Information Science and Technology*, *53*(1), 1-4. https://doi.org/10.1002/pra2.2016.14505301098

Jacsó, P. (2012). Google Scholar Author Citation Tracker: Is it too little, too late? *Online Information Review*, *36*(1), 126-141. https://doi.org/10.1108/14684521211209581

Kim, H. J. & Grofman, B. (2020). Who Creates a Google Scholar Profile? *PS: Political Science & Politics*, *53*(3), 515-520. https://doi.org/10.1017/S1049096520000189

Lammey, R. (2020). Solutions for identification problems: A look at the Research Organization Registry. *Science Editing*, *7*(1), 65-69. https://doi.org/10.6087/kcse.192

Martín-Martín, A., Orduna-Malea, E. & Delgado López-Cózar, E. (2018). A novel method for depicting academic disciplines through Google Scholar Citations: The case of Bibliometrics. *Scientometrics*, *114*(3), 1251-1273. https://doi.org/10.1007/s11192-017-2587-4

Meadows, A. (4 December 2019). Are You Ready to ROR? An Inside Look at this New Organization Identifier Registry. *The Scholarly kitchen* [blog]. https://scholarlykitchen.sspnet.org/2019/12/04/are-you-ready-to-ror-an-inside-look-at-this-new-organization-identifier-registry

Mikki, S., Zygmuntowska, M., Gjesdal, Ø. L. & Al Ruwehy, H. A. (2015). Digital Presence of Norwegian Scholars on Academic Network Sites—Where and Who Are They? *PLOS ONE*, *10*(11), e0142709. https://doi.org/10.1371/journal.pone.0142709

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A. & Delgado López-Cózar, E. (2017). The lost academic home: Institutional affiliation links in Google Scholar Citations. *Online Information Review*, *41*(6), 762-781. https://doi.org/10.1108/OIR-10-2016-0302

Orduña-Malea, E., & Bautista-Puig, N. (2022). Measuring web connectivity between research organizations through ROR identifiers. *26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*, Granada, Spain. Zenodo. https://doi.org/10.5281/zenodo.6948453

Ortega, J. L., & Aguillo, I. F. (2013). Institutional and country collaboration in an online service of scientific profiles: Google Scholar Citations. *Journal of Informetrics*, *7*(2), 394-403. https://doi.org/10.1016/j.joi.2012.12.007

Ortega, J. L. (2015a). Diferencias y evolución del impacto académico en los perfiles de Google Scholar Citations: Una aplicación de árboles de decisión. *Revista española de Documentación Científica*, *38*(4), e102. https://doi.org/10.3989/redc.2015.4.1225

Ortega, J. L. (2015b). How is an academic social site populated? A demographic study of Google Scholar Citations population. *Scientometrics*, *104*(1), 1-18. https://doi.org/10.1007/s11192-015-1593-7

Ortega, J. L. (2015c). Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. *Journal of Informetrics*, *9*(1), 39-49. https://doi.org/10.1016/j.joi.2014.11.004

Ortega, J. L. & Aguillo, I. F. (2012). Science is all in the eye of the beholder: Keyword maps in Google scholar citations. *Journal of the American Society for Information Science and Technology*, *63*(12), 2370-2377. https://doi.org/10.1002/asi.22761

Ortega, J. L. & Aguillo, I. F. (2013). Institutional and country collaboration in an online service of scientific profiles: Google Scholar Citations. *Journal of Informetrics*, 7(2), 394-403. https://doi.org/10.1016/j.joi.2012.12.007

Ortega, J. L. & Aguillo, I. F. (2014). Microsoft academic search and Google scholar citations: Comparative analysis of author profiles: Microsoft Academic Search and Google Scholar Citations: Comparative Analysis of Author Profiles. *Journal of the Association for Information Science and Technology*, 65(6), 1149-1156. https://doi.org/10.1002/asi.23036

Teixeira da Silva, J. A. (2018). The Google Scholar h-index: Useful but burdensome metric. *Scientometrics*, 117(1), 631-635. https://doi.org/10.1007/s11192-018-2859-7

Teixeira da Silva, J. A. (2021). The i100-index, i1000-index and i10,000-index: Expansion and fortification of the Google Scholar h-index for finer-scale citation descriptions and researcher classification. *Scientometrics*, 126(4), 3667-3672. https://doi.org/10.1007/s11192-020-03831-9

Thoma, B. & Chan, T. M. (2019). Using Google Scholar to track the scholarly output of research groups. *Perspectives on Medical Education*, 8(3), 201-205. https://doi.org/10.1007/s40037-019-0515-4

Tsou, A., Bowman, T. D., Sugimoto, T., Lariviere, V. & Sugimoto, C. R. (2016). Self-presentation in scholarly profiles: Characteristics of images and perceptions of professionalism and attractiveness on academic social networking sites. *First Monday*, 21(4). http://dx.doi.org/10.5210/fm.v21i4.6381

CV

Enrique Orduña-Malea. He is Technical Telecommunication Engineer, MA in Library Science, Master in Management of Multichannel Contents, and holds a PhD with a dissertation thesis about Web university rankings (Spanish thesis award). Currently, he works as a tenured Associate Professor in the Department of Audiovisual Communication, Documentation and History of Art, at the Universitat Politècnica de València. His lines of research are mainly related to the application of web-based techniques to Science Studies. Enrique also collaborates in funded research projects and international advisory activities about the web visibility of universities, researchers, journalsand other online research objects.

Núria Bautista-Puig. She attained her BsC in Geography (Universitat de Lleida, Spain), MSc in GIS and remote sensing (Universidad de Zaragoza, Spain) and Ph.D. in Library and Information Science (2020, Universidad Carlos III de Madrid, AASHE Campus Sustainability Research Award, Spanish thesis award). Currently, she is an Assistant Professor at the Departament of Librarianship and Documentation at the Universidad Complutense de Madrid. Her areas of expertise include the fields of information science, organizational sustainability and bibliometrics/scientometrics.