Exploring ORCID adoption and metadata presence in Spain's research landscape

Wenceslao Arroyo-Machado[1*], Benjamín Vargas-Quesada[2], Teresa Muñoz-Écija[3] and Zaida Chinchilla-Rodríguez[4]

warroyom@asu.edu; benjamin@ugr.es; teresamunyozecija@gmail.com; zaida.chinchilla@csic.es
0000-0001-9437-8757; 0000-0001-5115-7460; 0000-0003-1311-0471; 0000-0002-1608-4478

[1]Center for Science, Technology and Environmental Policy Studies, School of Public Affairs, Arizona State University, Phoenix, AZ 85004, USA
[2]Unit for Computational Humanities and Social Sciences (U-CHASS), University of Granada, Granada, Spain
[3]Independent researcher, Granada, Spain
[4]Institute of Public Goods and Policies (IPP), CSIC, Madrid, Spain

*Corresponding author

Abstract

In the evolving landscape of scientific research, ORCID identifiers are essential for maintaining academic integrity and enhancing research visibility. This study proposes a methodological framework for identifying and analyzing the presence and characteristics of researchers in a specific country in ORCID, using complementary data from OpenAlex to improve results. Our objectives were to develop a reliable method to identify ORCID records of researchers in Spain, study the frequency of profile updates, and investigate the most frequently filled metadata sections and common profile characteristics across disciplines. We identified 190,455 ORCID records of Spanish researchers and found high engagement with some metadata, particularly in the 'Employment' (73%) and 'Publications' (83%) sections. Additionally, 73% of records were actively maintained. The most common metadata combination included 'Works', 'Employment', and 'Education and Qualifications'. However, we noted significant variability and incomplete records across disciplines. The study highlights the need for more complete and frequently updated profiles, improved institutional integration, and strategic efforts to enhance ORCID adoption. These improvements could enhance the reliability of ORCID for tracking researcher mobility and academic trajectories.

Keywords: ORCID; OpenAlex; scientometrics; Spain; researcher profiles

Key points

- A methodological framework has been proposed for accurately identifying researchers from a specific country within the ORCID database.
- 73% of the 190,455 ORCID records of Spanish researchers are active.
- Key fields like 'Employment' and 'Works' are highly populated, while 'Keywords' and 'Peer review' are less frequently filled.
- Common metadata combinations often include 'Works', 'Employment', 'Education and Qualifications', and 'Other IDs'.
- The majority of ORCID records fall under Physical Sciences followed by Social Sciences, Health Sciences, and Life Sciences.

1. INTRODUCTION

In the current dynamic and global context of ever-changing scientific research, persistent identifiers (PIDs) have become an important tool for researchers, facilitating reliable discovery, citation, and interlinking of digital resources over time. PIDs address the complexities introduced by shifting locations and access methods to information (Meadows et al., 2019). Adopting these identifiers helps track scholarly patterns and keep academic integrity and enhances knowledge networks, proving essential for the scholarly system (Macgregor et al., 2023). By enabling seamless interconnection, interoperability of information, and cross-disciplinary collaborations, PIDs are crucial for maintaining the coherence and continuity of the open science ecosystem, thereby bolstering its core principles of transparency, accessibility, and scientific collaboration (Anglada & Abadal, 2018).

The shift towards an open science paradigm needs research infrastructures that provide open scholarly metadata, enhancing the visibility, openness, aggregation, and management of research outputs (Bornmann et al., 2021; Peroni & Shotton, 2020). In this context, ORCID (Open Researcher and Contributor Identifier) has become pivotal. ORCID stands out by offering a unique and persistent identification system that bridges authors with their diverse contributions across the academic and research landscapes. It achieves this through ORCID IDs, which not only link researchers to their works, activities, and affiliations but also help these contributions be accurately recognized and transparently managed in the scholarly ecosystem (ORCID, 2024a).

The integration of ORCID thus represents a significant advance in the organization of the academic landscape, promoting transparency in the registration and acknowledged of researchers' efforts. Nevertheless, it has been posited that the motivation of researchers to engage with ORCID is primarily driven by mandates from publishing and funding entities within the research ecosystem (Porter, 2022; Teixeira da Silva, 2021b) rather than the intrinsic advantages associated with maintaining an ORCID profile. This perspective emphasises a compliance-driven adoption and highlight the need to effectively communicate and leverage the intrinsic value of ORCID profiles to researchers themselves. This may be attributed to the complexity of integrating ORCID into existing research infrastructures (Dappert et al., 2017; Schnieders et al., 2022).

Although the adoption of ORCID as a PID for unique author identification is becoming increasingly common (Haak et al., 2012), there are several challenges that researchers face when working with ORCID as a data source for scientometric studies. For example, there is a significant lack of adoption and completeness in ORCID records that varies across disciplines (Fernández-Marcial et al., 2023), the hard sciences tend to lead in adoption and completeness, in contrast to the social sciences, arts, and humanities (Boudry & Durand-Barthez, 2020; Bordons et al., 2024), and also across countries which leads to biases in coverage (Youtie et al., 2017). Previous studies have identified quality data issues such as empty records, problems with homonyms, sections with outdated data, multiple profiles for the same author (Heusse & Cabanac, 2022; Wang et al., 2024); the

occurrence of duplicate and fake profiles (Baglioni, Mannocci, et al., 2021; Teixeira da Silva, 2021a; 2021c), the creation of profiles by personnel not directly linked to research (Heusse & Cabanac, 2022), regular update of profiles (Costas et al., 2022), misidentification of authors and their associated institutions (Martínez-Méndez & Lopez-Carreño, 2019), the creation of "silent" or 'ghost' profiles (Teixeira da Silva, 2021a; 2023; Wang et al., 2024) and the limitations in data verification (Wang et al., 2024). Furthermore, there is a lack of standardization of information, which often reflects a compliance response to mandates from both publishers and funders (Chorąś & Jaroszewska-Chorąś, 2020; Porter, 2022; Teixeira da Silva, 2022).

Despite these challenges, researchers generally have a positive perception of PIDs and their scientific infrastructure. However, there is a pressing need for user training to avoid inconsistencies and errors (Houghton & Foster, 2024), and dissemination efforts are crucial to address the evident unawareness about their identification, purpose, and utility. This lack of knowledge varies across disciplines and scholarly occupations (Macgregor et al., 2023). In this line, the importance of developing and implementing projects by research institutions, along with promotional campaigns, has been identified as crucial tools for successful ORCID adoption and use within the research community (Aghassibake et al., 2023; Pampel et al., 2024).

To date, there has been a lack of prior systematic studies investigating these phenomena in Spain. This paper aims to fill this gap in the literature by comprehensively analysing the adoption and use of ORCID identifiers and metadata completeness for Spanish researchers. To address this research, the following specific objectives have been established:

1. To develop a methodology for accurately identifying Spanish researchers within the ORCID database, providing a reliable base for further analysis of the adoption rate and data completeness.
2. To investigate the metadata sections most frequently updated and completed by Spanish researchers in their ORCID profiles.
3. To explore the differences across disciplines to observe if there are any discernible patterns in this regard.

We hope that this study will provide insights into the use and practices of ORCID in the Spanish scholarly ecosystem and may serve to inspire further research that helps to enhance the understanding of scholars in the adoption of ORCID in their work and the compromise of scientific and academic organizations in facilitating training and infrastructure to its adoption.

1.1. Literature review

ORCID is considered a data source that can facilitate meta-research by enhancing the visibility of a wider array of scientific activities and promoting open science practices (Costas et al., 2022), thus contributing to the FAIR principles (Wilkinson et al., 2016). This potential use of ORCID has been revealed as fundamental in advancement and

recognition within the scientific field, highlighting their importance in documenting and making these aspects visible. Moreover, ORCID is particularly renowned for its support of author identifiers and its interoperability with other databases and systems (Velez-Estevez et al., 2023).

In examining the integration of ORCID with other databases, studies have demonstrated its efficacy in elucidating the dynamics of scientific research. For example, the prevalence of ORCID records in databases such as PubMed and MEDLINE is more pronounced in systems where metadata is collected automatically rather than requiring manual input by authors (Boudry, 2021). Additional research highlights successful integrations of ORCID within other data sources. For example, CrossRef and PubMed have incorporated ORCID identifiers to streamline author identification and enhance the accuracy of publication records (Boudry, 2021; Haak et al., 2012). An analysis using the Dimensions database found that different regions adopted and actively used ORCID IDs at different rates. Portugal and Australia seemed to be ahead of the curve. This difference might be because some funding agencies require ORCID IDs, and the importance of ORCID IDs also likely varies depending on the specific field of research (Porter, 2022).

DataCite, another metadata source, has been examined for its linkage with research institutions. Findings indicate an increase in best practices such as incorporating ORCID identifiers. Nevertheless, there is a need for improved connections between data and associated publications (Van Wettere, 2021). Recognizing these advancements in enhancing metadata's role in research, ORCID's contributions are not limited to metadata improvement alone. ORCID's role extends to facilitating scientific evaluation processes, mapping data sources that were previously difficult to access, and supporting equitable and representative evaluation of research contributions (Haak et al., 2018).

Despite these advancements in metadata improvement and scientific evaluation, the adoption and perception of ORCID vary significantly across different regions and institutions. In Toulouse's scientific complex, 41.8% of personnel have adopted ORCID, but its use varies among disciplines such as Health and Economics, revealing a limited understanding and utilization of this identifier (Heusse & Cabanac, 2022). Similarly, a study of Irish faculty members at the Technological Universities and Institutes of Technology in Ireland found that while more than two-thirds have an ORCID profile, it is mainly used for promoting their work. Notably, there is a significant negative perception towards ORCID, with concerns about its purpose, external pressure to register, (from universities, journals, and funders), fears of being monitored (concerns over inaccuracies in tracking academic or teaching activities), reluctance to keep profiles updated, and worries over security and privacy vulnerabilities (Houghton & Foster, 2024).

Within the university framework, a range of practices and perceptions are observed, as universities are integrating ORCID identifiers into their institutional repositories and researcher profile systems to improve data management and ensure consistent researcher

profiles (Thomas et al., 2015). While at the University of Caen in Normandy, a study by Boudry and Durand-Barthez (2020) analyzed the presence of researchers across several academic networking platforms, including ORCID, ResearcherID, Academia.edu, and ResearchGate. The findings revealed that 64.3% of researchers had profiles on both ORCID and ResearchGate, with ResearchGate being the more widely used platform. Despite this, the study highlighted a relatively low adoption rate of ORCID and ResearcherID, compounded by the absence of comprehensive publication references within these profiles. This lack of detailed bibliometric information poses significant challenges for accurate bibliometric evaluation and the broader aim of promoting open science practices. In Portugal, the Faculty of Arts and Humanities at the University of Porto showed a high ORCID adoption rate (90.4%), but the records were often incomplete, impacting the disambiguation of authors. This study emphasized inconsistencies in affiliation information, underscoring the need for institutional support and interoperability to ensure the completeness and accuracy of profiles (Fernández-Marcial et al., 2023). Similarly, at Emory University's School of Law, the challenges faced in promoting the adoption of ORCID among legal researchers were identified, given that their publishing and communication models are more traditional than in the sciences. Continuous training and dissemination were highlighted as essential solutions for the ideal adoption of this persistent identifier (Quinn, 2023).

Similarly, since its introduction in Spanish institutions in 2013, initiated by the University of Oviedo and backed by the FECYT, ORCID's integration has signified a pivotal shift towards enhancing interoperability within research infrastructures. This move reflects a strong dedication to refining research methodologies and governance of ORCID, thereby bolstering its application across Spain's scientific landscape (Marín-Arraiza & Mejias, 2020). This evolution resonates with key principles and frameworks such as the DORA principles (American Society for Cell Biology, n.d.), the Leiden Manifesto (Hicks et al., 2015), CoARA (2022), the recent Spanish university law ("Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario," 2023), the National Strategy for Open Science 2023-2027 (Ministry of Science and Innovation, 2023), and the Barcelona Declaration on Open Research Information (2024).

ORCID's alignment with these initiatives underscores its crucial role in reinforcing standardization and academic recognition, promoting integrity, collaboration, and the dissemination of knowledge within the research community (Marín-Arraiza & Mejias, 2020). The current Spanish scenario provides the opportunity to reflect on the use of PIDs and explore optimal strategies for their implementation or integration into existing open science processes (Marín-Arraiza, 2022), addressing metadata accessibility issues and further consolidating the framework of an open information ecosystem (Delgado-López-Cozar & Martín-Martín, 2024).

In light of the broader discussions on the strategic use of PIDs, it is important to examine specific studies that illustrate ORCID adoption within the Spanish research community. To date, we know that in Spain, found that the ORCID identifier is included in most

articles published in Q1 and Q2 Spanish open access journals, with a predominant presence in Q1 journals, and that the Spanish National Research Council (CSIC) is the leading Spanish scientific publisher that has adopted it in its editorial management, suggesting that the ORCID identifier could become a criterion for editorial quality (Martínez Méndez & López Carreño, 2019). Bordons et al. (2024) analyzed ORCID adoption in Spanish scientific articles and among ERC grant recipients in Spain. They found that almost 90% of articles list at least one ORCID ID, but only 14% include ORCID IDs for all authors. All ERC grantees had ORCID IDs, with 50% updating their profiles monthly. Differences by gender, discipline, and funding scheme were also examined, insisting that publisher and journal policies are essential to drive ORCID adoption.

2. METHODOLOGY

2.1. Dataset generation

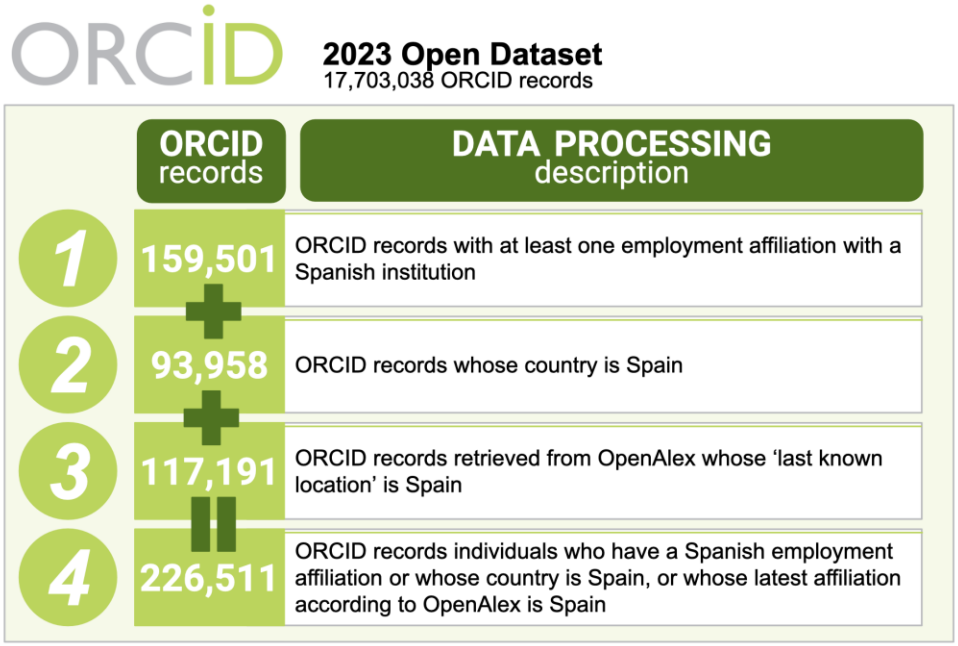
We utilized ORCID and OpenAlex as open data sources for our analysis. From ORCID's public annual snapshot of September 2023, we collected data on over 17 million profiles. These profiles provided a rich set of metadata, including identification details, employment affiliations, educational background, publication lists, and peer review contributions, as summarized in Table 1. It also provides information about metadata update logs. Additionally, we accessed data from OpenAlex, updated to April 2024, through the InSySPo cluster to enrich our analysis by retrieving additional ORCID profiles and uniquely identifying researchers and their main research areas.

Table 1. ORCID record metadata analyzed

Section	Field	Description
Profile identification data	Name	Full name
	Other names	Name variants
	Biography	Brief professional career description
	Email(s)	Email address(es)
	Websites and social media links	Links to websites and social media profiles
	Other IDs	Researcher identifiers like ISNI, Scopus Author ID, ResearcherID
	Keywords	Words describing research activity
	Countries	Country of professional activity
Activities	Employment	Information about professional affiliations with organizations
	Education and qualifications	Details on educational institutions and degrees or professional qualifications
	Professional activities	Details about affiliations with various organizations, societies, or associations, and roles in supporting these entities
	Funding	Details on grants, awards, and other types of research support funding
	Works	Publications, books, conferences, and other academic contributions
Additional feature	Peer review	Allows users to link and display their contributions as reviewers on their ORCID profile, documenting their participation in peer review processes

To construct the sample, we conducted three searches to collect our dataset. First, from the ORCID public snapshot of September 2023, we collected records listing a Spanish affiliation in the 'Employment' field. Second, we included records from the same ORCID snapshot where the 'Country' field was listed as Spain. Third, we used the OpenAlex public snapshot, updated as of April 2024 and accessed through the InSySPo cluster, to identify additional ORCID records of Spanish researchers that were not retrievable by the previous methods. We retrieved records of Spanish authors whose last known affiliation is a Spanish institution and filtered authors having an ORCID ID (15.88%). This last filter allowed us to identify 19% of records (44,055) that have no affiliation or country associated with them, making it impossible to locate using ORCID alone. These three searches were combined to avoid duplication, resulting in a dataset of 226,511 records. Figure 1 illustrates the steps followed for retrieving the Spanish ORCID records.

Figure 1. Summary of Spanish ORCID records retrieval process



2.2. Researchers and main research area identification

To determine the use of ORCID in Spain, and to exclude non-researcher profiles such as librarians, technical staff, etc., we carried out a procedure to identify researchers as comprehensively as possible among all individuals in our dataset. It should be noted that this process does not attempt to identify precisely the different academic positions or to limit ourselves to those within the university or research centre, but rather to identify all those who have carried out research activities, and to avoid including profiles that are significantly different. First, we used the last self-declared role in their ORCID record in the Employment field. In total, 139,335 (61.5%) of the ORCID profiles have a role listed. However, it lacks standardization, with 43,086 unique roles. Manually, this field was cleaned using basic text mining techniques, eliminating special characters, transforming text strings to lowercase, and correcting errors such as double spaces. However, this reduced the number of roles to 38,250. Therefore, the 2,794 most common roles, which appear in 70% of the ORCID records that include this field, were manually reviewed to

identify which were academic roles and which were not. Subsequently, matches were sought for terms marked as researcher profiles among the remaining 30% of roles. In total, 23,678 roles were identified as academic, thus marking 77% of ORCID users in this field as researchers, representing 47% of the general dataset. To this process, we have added the tagging as a researcher of all ORCID records present in OpenAlex, as well as all ORCID records with a Web of Science or Scopus identifier or more than 4 journal publications. It should be noted that there may be discrepancies between the different processes, but the approach is to be as inclusive as possible and to have the broadest possible representation of researchers. In total, for the complete dataset, 190,455 (80%) were marked as researchers, and these are the records to be used in this study.

On the other hand, for the identification of the main research area of each ORCID record, data from both ORCID and OpenAlex were utilized. First, we retrieved all DOIs (Digital Object Identifiers) of the publications listed in each ORCID profile. Next, using these DOIs, we located all corresponding publications indexed in the OpenAlex database. Additionally, for all authors indexed in OpenAlex with an ORCID identifier, we retrieved all their publications from OpenAlex. Secondly, for all ORCID records present in OpenAlex, all their publications were retrieved. After that, using the OpenAlex classification system of citation topics developed by CWTS, a count of the domains associated with each publication was carried out. This is the highest hierarchical level of the topic classification and consists of the 4 main areas (Life Sciences, Social Sciences, Physical Sciences, and Health Sciences). In this way, each ORCID record was associated with a domain in which it had the most publications, with the possibility of a tie in some cases.

2.3. Methods

For the data analysis, various statistical methods were applied using R. Initially, descriptive statistics were employed to explore the characteristics, activity, and age of ORCID records for Spanish researchers, as well as the metadata they complete. This initial analysis provided an overview of the dataset, highlighting key trends and patterns in the information provided by researchers. To gain a more detailed understanding, we expanded this study by analyzing the intersections of fields within these records, utilizing UpSet plots. These plots allowed us to visualize the combinations of metadata fields that are most frequently filled, offering insights into how researchers populate their profiles. All data and scripts used for this analysis are available on Zenodo (10.5281/zenodo.12019457) and GitHub (https://github.com/Wences91/spanish_orcids/tree/main).

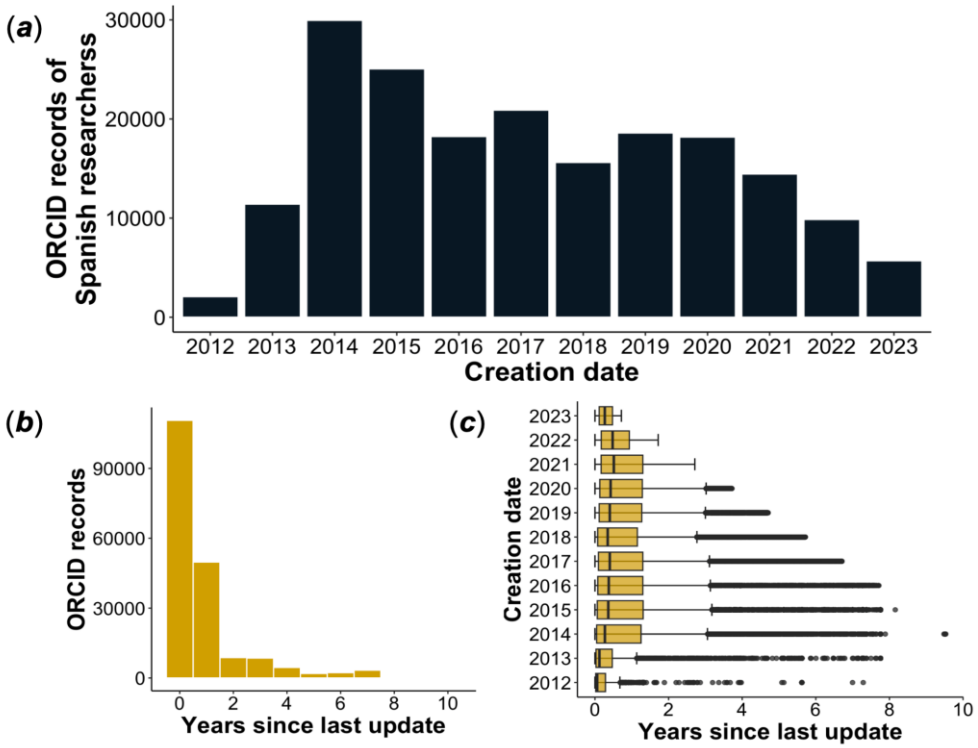
3. RESULTS

3.1. Activity of Spanish researchers in ORCID

There are 190,455 Spanish researchers registered in ORCID, with 139,004 (73%) being active accounts, indicating recent activity within the last year (September 2022 to September 2023). The data shows a significant increase in profile creation starting in 2012, with a marked peak between 2014 and 2017 (Figure 2a). This period saw the highest

number of registrations, a total of 94,152 new records (49%). The adoption of ORCID profiles in Spain has shown a significant upward trend over recent years. This surge can be attributed to several factors, primarily the increasing mandates from academic institutions and scholarly journals. Universities and research organizations in Spain have progressively required their faculty and researchers to register for ORCID iDs to ensure accurate attribution of their work and to streamline administrative processes. Additionally, many academic journals now mandate ORCID iDs for authors as part of the submission process, aiming to enhance the integrity and traceability of scholarly publications. These institutional and editorial requirements have collectively driven a substantial increase in ORCID registrations, resulting in a notable peak in the number of profiles created. Following this peak, there is a noticeable decline in new ORCID profile creations, continuing through 2023. However, it is important to note that the ORCID data only covers up to September 2023. This trend suggests a stabilization in the adoption of ORCID among Spanish researchers, possibly indicating that most eligible researchers have already registered. The decrease in new registrations could also reflect a maturation of the ORCID system within Spain. Initially, the growth was rapid due to increased awareness and the implementation of mandatory policies, whereas the current growth rate is more aligned with the natural influx of new researchers entering the system. Additional factors contributing to the decline may include the saturation of the target population, reduced promotional efforts, administrative challenges, a perceived lack of utility by some researchers, and the use of alternative identification systems (ORCID, 2024b).

Figure 2. Distribution of Spanish ORCID records by a) creation date, b) years since last update, c) and years since last update per creation date



The distribution of Spanish ORCID profiles based on the years since their last update reveals that the majority have been updated within the past year (58%), highlighting a high level of recent activity and engagement (Figure 2b). The frequency of profile updates declines sharply after one year, with a significant drop in profiles not updated for more than one year (16%). The analysis of update frequency, categorized by the year of profile creation, shows that the median time since the last update remains around one year across all creation years (Figure 2c). This regular updating behavior reflects an active academic community that maintains their profiles to keep their information current despite the ORCID record age. While a core group of researchers remains highly active, there are segments of the research community that update less frequently, which could reflect diverse career stages and varying degrees of ongoing research activity.

The observed high frequency of recent updates may also be influenced by institutional infrastructures that automatically update researchers' profiles. Many universities and research institutions integrate ORCID with their internal systems, ensuring that profiles are kept up to date with minimal effort from individual researchers. This automation supports continuous and consistent profile management, contributing to the overall activity levels observed.

However, these data primarily indicate that most profiles show a certain degree of activity, regardless of the age of the registration. They do not, however, clearly reflect the extent to which these updates maintain the profiles' completeness and accuracy. In the following section, the level of completeness of the profiles will be examined in more detail to better understand the adoption and usage of this platform by Spanish researchers.

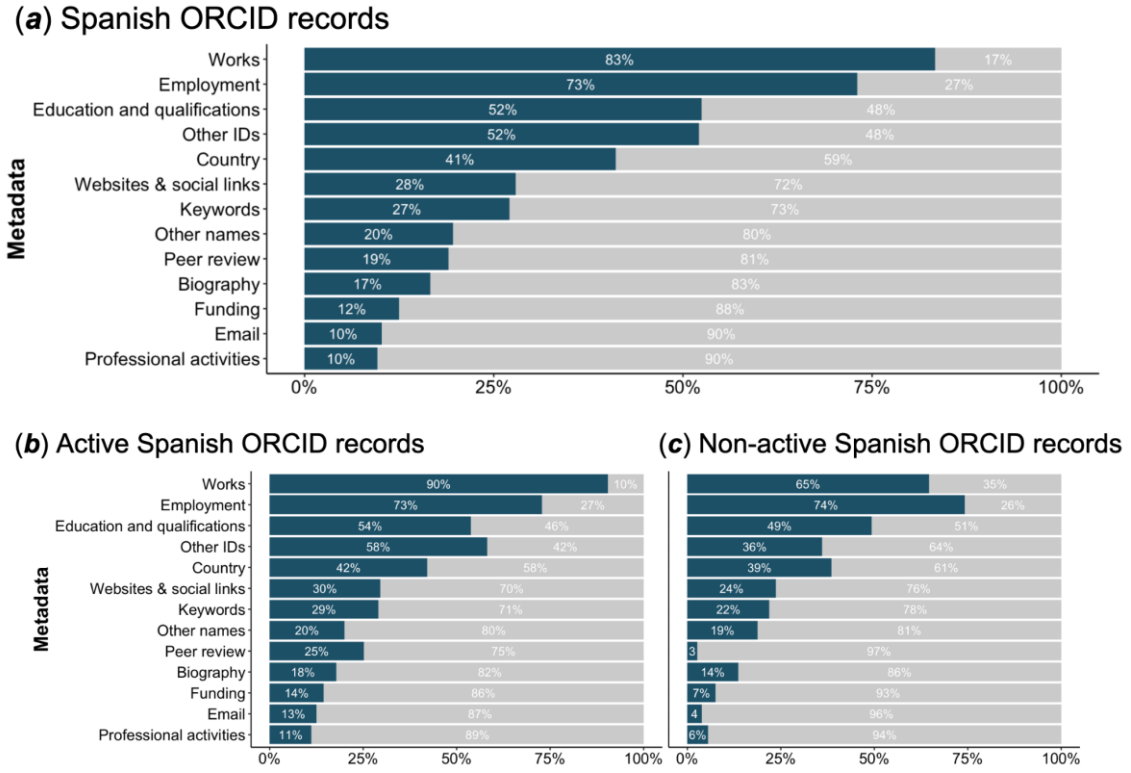
3.2. Metadata analysis

The metadata analysis of Spanish ORCID records (Figure 3a) reveals notable disparities in the completeness of various sections. Overall, 83% of the records include information in the ‘Works’ section and 73% in the ‘Employment’ section. However, other sections, such as ‘Keywords’ (27%), ‘Other names’ (20%), and ‘Peer review’ (19%), are less frequently completed. Of the 13 fields studied, only 4 reflect values above 50%. This disparity can stem from multiple factors such as a lack of activity in specific areas (e.g., professional activities or funding grants) or insufficient information to provide in certain fields (e.g., alternative names). Privacy concerns (e.g., email or address) or lack of connections to other platforms (e.g., peer review or other IDs) can also contribute to these differences. In cases such as ‘Keywords’, the lower levels of information could be due to a lack of interest or lack of knowledge on the part of the researchers themselves.

In contrast, the analysis of active and non-active Spanish ORCID records provides deeper insights into these information presence trends. Active records show a significantly higher presence of information in the main sections, specifically 90% for ‘Works’ and 73% for ‘Employment’. Active profiles also show moderate levels of information in ‘Education and Qualifications’ (54%) and ‘Other IDs’ (58%). However, there is less information included in sections such as ‘Keywords’ (29%) and ‘Peer review’ (25%). On the other

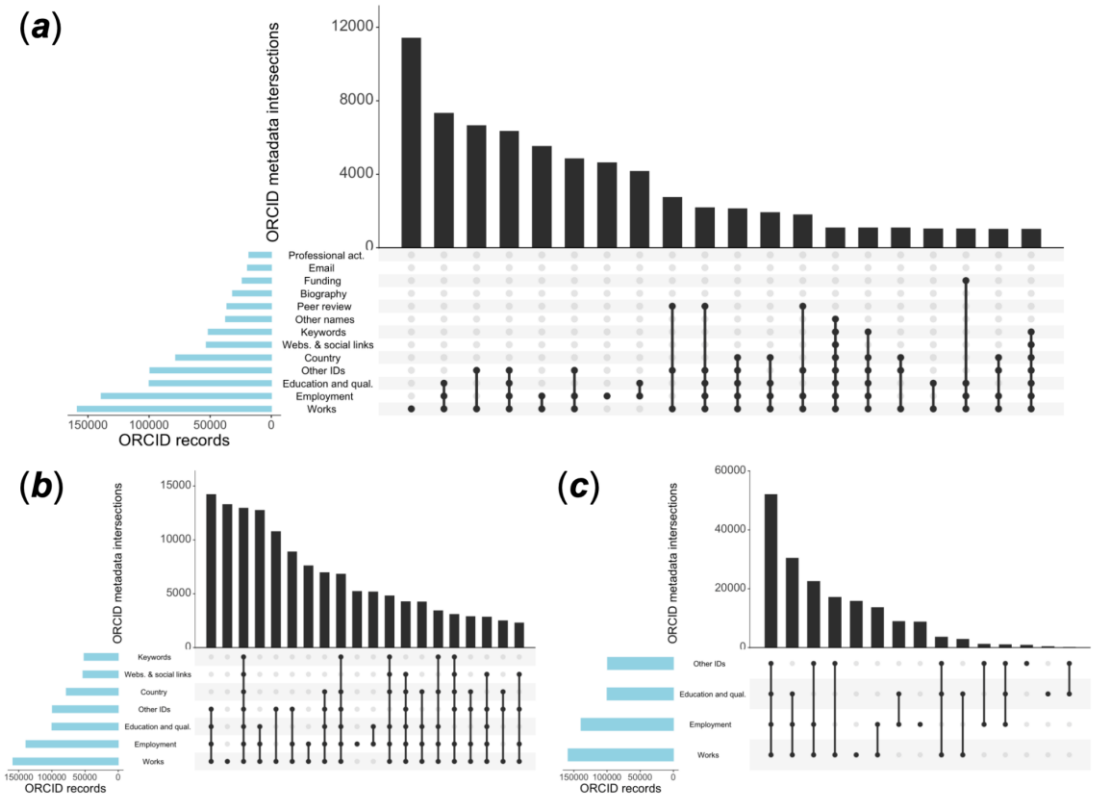
hand, non-active records have noticeably lower levels of information across most sections. For instance, only 65% of non-active profiles have data included in the 'Works' section, and 49% have information in the 'Employment' section. The low levels of information for sections such as 'Email' (7%) and 'Professional Activities' (6%) in non-active profiles highlight the broader issue of outdated or minimally maintained records. This gap between active and non-active profiles suggests that many researchers may create ORCID accounts primarily for initial use and then neglect to update them regularly, leading to incomplete and less useful records.

Figure 3. Availability of metadata in Spanish ORCID researcher records: a) all records; b) active records; c) non-active records



Beyond the percentage of records that include a specific field, it is essential to consider how these fields appear together. Among all possible combinations, the 20 most frequent cover 69,251 records (36%), with a total of 8,178 possible combinations. The most common case is the record that only includes 'Works', with 11,423 records (6%), followed by those that include 'Works', 'Employment', and 'Education and Qualifications' (4%) (Figure 4a). This is expected, given that 6 of the 13 analyzed fields appear in only 20% or fewer records. When this analysis excludes these less-used fields, the most frequent case is the one that includes 'Works', 'Employment', 'Education and Qualifications', and 'Other IDs' (7%) (Figure 4b). When selecting only fields present in at least 50% of the records (Figure 4c), the most frequent case again involves all these fields (27%). This underscores the complexity of accounting for fields and profiles, ultimately reflecting a broad diversity of profiles. Despite this, there does appear to be a dominant pattern, particularly among records that include all four fields.

Figure 4. Top 20 metadata intersections of Spanish ORCID records considering: a) all ORCID fields, b) only ORCID fields with a prevalence of more than 20%, and c) only ORCID fields with a prevalence of more than 50%



3.3. Differences across research domains

Researchers were classified into four major areas of knowledge or domains according to OpenAlex (Table 2): Life Sciences, Social Sciences, Physical Sciences, and Health Sciences. The majority of ORCID records fall under Physical Sciences (37.9%), followed by Social Sciences (26.3%), Health Sciences (24.6%), and Life Sciences (18%). Notably, over 80% of the records in each domain are active, indicating strong engagement from researchers regardless of the domain. The average number of journal publications is highest in the Physical Sciences (31.4) and Health Sciences (28.3). There is a slight bias because classifying ORCID profiles by domain requires at least one publication, but even with this bias, the average number of journal publications remains high at 25.9. Additionally, there are epistemic differences across the sciences, with Social Sciences having the lowest average number of publications (15.1), which, although lower, is still relatively high. In the end, what this elevated average indicates is that researchers who provide publication data tend to have a higher volume of publications.

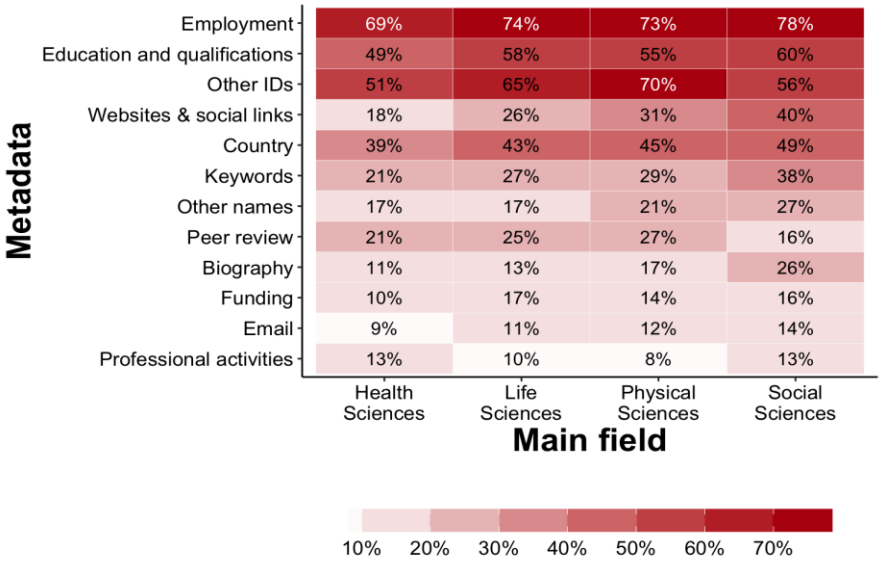
Table 2. Distribution of Spanish ORCID records across research domains and average publications

	Life Sciences	Social Sciences	Physical Sciences	Health Sciences	Total
ORCID records	26,864 (18%)	39,277 (26.3%)	56,513 (37.9%)	36,638 (24.6%)	149,182 (100%)
Active records	21,629 (80.5%)	31,356 (79.8%)	45,644 (80.8%)	30,073 (82.1%)	122,836 (82.3%)
Avg. Works	30,1 ±47.9	25.6 ±42.4	40.6 ±86	30.9 ±68.3	32.9 ±67.5
Avg. Journal publications	27 ±44	15.1 ±24	31.4 ±75.1	28.3 ±59.8	25.9 ±57.7

Note: 'Works' refers to the total elements whereas 'Journal publications' are the works published in journals.

There are differences and commonalities in how Spanish researchers across different domains populate their ORCID profiles (Figure 5). Employment information is prominently featured across all fields, with notable percentages such as 69% in Health Sciences and 78% in Social Sciences. Similarly, fields like 'Education and qualifications' and 'Other IDs' also show high percentages, especially in domains like Life Sciences and Physical Sciences, reflecting an emphasis or predominance of multiple academic identifiers. In contrast, less emphasis is placed on fields such as 'Keywords' and 'Peer review', with 'Keywords' appearing most frequently in Social Sciences at 38% and 'Peer review' in Physical Sciences at 27%. While some metadata fields are universally prioritized, others are more domain-specific. In this sense, Social Sciences consistently show higher percentages across multiple fields, indicating a more comprehensive approach to metadata entry compared to other domains.

Figure 5. Availability of metadata in Spanish ORCID researcher records across research domains



4. DISCUSSION

Our study reveals a strong representation of Spanish researchers in ORCID, especially among those actively engaged in research, providing a clear response to the inquiry regarding their ORCID involvement (Porter, 2022). It is evident from the metadata analysis that 'Employment' and 'Publications' are the sections most frequently populated by Spanish researchers in their ORCID profiles, emphasizing the importance placed on career affiliation and scholarly outputs (Youtie et al., 2017). However, not all metadata sections in an ORCID record are completed exhaustively, leading to profiles that are partially filled (Fernández-Marcial et al., 2023; Heusse & Cabanac, 2022), which may inadvertently result in the undervaluation of other academic contributions and activities.

The completeness of metadata sections in ORCID records varies, and the presence of partially filled profiles could pose challenges to their application in science of science research (Thompson et al., 2022). Incomplete records may skew the data and lead to inaccuracies in the analysis, affecting the validity of studies that rely on this information. This issue of data integrity becomes critical when considering ORCID as an open metadata source for science of science research (Costas et al., 2022).

The principle of interoperability, when applied to the automatic updating of ORCID profiles by various stakeholders within the research ecosystem, could markedly ameliorate the issue of incomplete metadata. By fostering an interoperable framework where research institutions, publishers, funding bodies, and ORCID seamlessly exchange data, we can ensure that a researcher's profile is consistently updated with the latest publications, grants, and affiliations. This level of integration would significantly improve the accuracy and completeness of ORCID records, enhancing its reliability as a resource for science studies (Baglioni et al., 2021). Interoperability facilitates not only the ease of data sharing across different platforms but also minimizes the manual effort required from researchers to maintain their profiles. Consequently, ORCID emerges as a more reliable, comprehensive open resource for documenting academic achievements, strengthening the foundation for in-depth analysis in science of science research.

Despite the benefits of name disambiguation (Sixto-Costoya et al., 2021), ORCID's detailed records could play a pivotal role in addressing and preventing authorship fraud. While our study has underlined the robust adoption of ORCID identifiers among Spanish researchers, it also ventures into the realm of academic integrity, demonstrating how the platform's capacity for detailed record-keeping aids in the combat against fraudulent practices (Baglioni et al., 2021; Teixeira da Silva, 2023).

By linking ORCID profiles of authors with Spanish affiliations, both active and inactive, to the OpenAlex database, our investigation gains a nuanced understanding of engagement levels with the ORCID system. This comprehensive approach reveals a landscape of ORCID engagement among Spanish researchers that is both committed and variable. Coupled with our calculations of the average profile update frequency, this

enriched data set provides a solid foundation for identifying discrepancies that may indicate academic misconduct (Khezr & Mohan, 2022; Teixeira da Silva, 2020). Our analysis suggests that the maintenance patterns of ORCID profiles, particularly the frequency of updates, can serve as indicators for potential 'paper mills' or phantom accounts (Teixeira da Silva, 2021a). Anomalies such as unusually high rates of profile updates or irregular co-authorship patterns, further highlighted by their OpenAlex associations, might signal fraudulent activities within the academic community.

Leveraging ORCID's comprehensive data, our study not only showcases the utility of ORCID within Spain but also emphasizes its potential to safeguard the research community from integrity threats. This presents an opportunity to develop a procedure for the Spanish institutional research system, guiding decision-makers, integrity research offices, and institutional ethical committees on best practices in scholarly communications and research. This approach aligns with methodologies to those proposed by Porter and McIntosh (2024), —enhanced by OpenAlex data—to improve transparency and traceability in scholarly contributions, thus strengthening the credibility of academic work.

Therefore, exploring the ramifications of ORCID's metadata richness is crucial not only in reflecting the active engagement of Spanish researchers but also as a potential asset in research ethics and integrity studies, as well as in the broader context of academic data ecosystems (Wouters et al., 2019).

5. CONCLUSIONS

This study provides a detailed overview of the adoption and use of ORCID by researchers in Spain. The results highlight a strong presence of active Spanish researcher profiles on ORCID, confirming the significance of this tool for managing academic identity and scientific visibility.

The uptake of ORCID among the Spanish academic community is robust, particularly among those actively engaged in research. The 'Employment' and 'Publications' sections are the most completed, highlighting the importance of professional affiliation and research outputs. Despite the high rate of active profiles, there is variability in the completeness of the metadata sections. This suggests that while ORCID is a valuable tool, its potential is not being fully realized due to partially completed and private profiles (Wang et al., 2024). Besides, as ORCID is recognized as potentially playing a crucial role in preventing authorship fraud (Teixeira da Silva, 2023), the interoperability with other databases and automatic updating of profiles by other entities in the research ecosystem (such as institutions, publishers, and funders) could significantly improve the completeness of metadata, thereby increasing the reliability of ORCID to carried out this type of studies (Teixeira da Silva, 2022; Bordons et al., 2024).

For future research, we plan to investigate the potential of ORCID as a tool for tracking researcher mobility and understanding academic trajectories. By analyzing ORCID records that include Spanish affiliations, we aim to conduct detailed mobility studies that

trace career progressions and geographic movements over time. This investigation will utilize the sequential order of affiliations within ORCID profiles, providing insights into the academic pathways and international flow of knowledge. Such research could illuminate patterns of researcher mobility, identify factors influencing academic career choices, and contribute to strategic planning in research policy and workforce development.

Several limitations should be considered in this research. The results of this research depend on the sources and the query used for data collection and comparison. On the one hand, although OpenAlex may be the most comprehensive database due to its open access (Visser et al., 2021), its integration with other data sources, and its support for open science and FAIR principles, it is possible that not all Spanish researchers are included. On the other hand, our sample could be affected by the OpenAlex author disambiguation, since cases of inconsistency have been detected (Culbert et al., 2024).

Acknowledgments

For further insights into the InSySPo infrastructure and associated projects, access the GitHub repositories at https://github.com/insyspo and https://github.com/alyssonmazoni.

REFERENCES

Aghassibake, N., Castello, O. G., Gujilde, P., & Rabun, S. (2023). Visualizing institutional activity using persistent identifier metadata. *Information Services & Use*, *Preprint*(Preprint), 1–8. https://doi.org/10.3233/ISU-230218

American Society for Cell Biology. (n.d.). *San Francisco Declaration on Research Assessment*. DORA. Retrieved March 24, 2024, from https://sfdora.org/read/

Anglada, L., & Abadal, E. (2018). ¿Qué es la ciencia abierta? *Anuario ThinkEPI*, *12*, 292. https://doi.org/10.3145/thinkepi.2018.43

Baglioni, M., Manghi, P., Mannocci, A., & Bardi, A. (2021). We Can Make a Better Use of ORCID: Five Observed Misapplications. *Data Science Journal*, *20*, 38. https://doi.org/10.5334/dsj-2021-038

Baglioni, M., Mannocci, A., Manghi, P., Atzori, C., Bardi, A., & Bruzzo, S. L. (2021). *Reflections on the Misuses of ORCID iDs*. https://doi.org/10.5281/ZENODO.4548308

Bornmann, L., Guns, R., Thelwall, M., & Wolfram, D. (2021). Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? An opinion paper. *Quantitative Science Studies*, *2*(2), 438–453. https://doi.org/10.1162/qss_e_00121

Barcelona Declaration on Open Research Information. (2024). Barcelona Declaration on Open Research Information. https://barcelona-declaration.org

Bordons, M., Moreno-Solano, L., & González-Albo, B. (2024). ORCID identifier adoption in Spanish scholarly communication: A macro and micro level perspective. *Learned Publishing*, *37*(3), https://doi.org/10.1002/leap.1606

Boudry, C. (2021). Availability of ORCIDs in publications archived in PubMed, MEDLINE, and Web of Science Core Collection. *Scientometrics*, *126*(4), 3355–3371. https://doi.org/10.1007/s11192-020-03825-7

Boudry, C., & Durand-Barthez, M. (2020). Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *PLOS ONE*, *15*(9), e0238583. https://doi.org/10.1371/journal.pone.0238583

Choraś, M., & Jaroszewska-Choraś, D. (2020). The scrutinizing look on the impending proliferation of mandatory ORCID use from the perspective of data protection, privacy and freedom of science. *Interdisciplinary Science Reviews*, *45*(4), 492–507. https://doi.org/10.1080/03080188.2020.1780773

CoARA (Coalition for Advancing Research Assessment). (2022). *Agreement on Reforming Research Assessment*. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

Costas, R., Corona, C., & Robinson-Garcia, N. (2022). *Could ORCID play a key role in meta-research? Discussing new analytical possibilities to study the dynamics of science and scientists* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/sjck6

Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2024). *Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus* (arXiv:2401.16359). arXiv. http://arxiv.org/abs/2401.16359

Dappert, A., Farquhar, A., Kotarski, R., & Hewlett, K. (2017). Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. *Data Science Journal*. https://doi.org/10.5334/dsj-2017-028

Delgado-López-Cozar, E., & Martín-Martín, A. (2024). La ruta de oro de la publicación científica: Del negocio de las revistas a las revistas negocio: La fuente del negocio editorial: El negocio bibliométrico de la evaluación científica. *Revista Mediterránea de Comunicación*. https://doi.org/10.14198/MEDCOM.26763

Fernández-Marcial, V., González-Solar, L., & Vale, A. (2023). Is ORCID your ID? A case study at the Faculty of Arts and Humanities of the University of Porto. *Learned Publishing*, *n/a*(n/a). https://doi.org/10.1002/leap.1562

Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*, *25*(4), 259–264. https://doi.org/10.1087/20120404

Haak, L. L., Meadows, A., & Brown, J. (2018). Using ORCID, DOI, and Other Open Identifiers in Research Evaluation. *Frontiers in Research Metrics and Analytics*, *3*, 28. https://doi.org/10.3389/frma.2018.00028

Heusse, M., & Cabanac, G. (2022). ORCID growth and field-wise dynamics of adoption: A case study of the Toulouse scientific area. *Learned Publishing*, *35*(4), 454–466. https://doi.org/10.1002/leap.1451

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Houghton, F., & Foster, A. (2024). Resistance and Power in Irish Higher Education: ORCID and the Monitored University. *The Journal of Academic Librarianship*, *50*(2), 102853. https://doi.org/10.1016/j.acalib.2024.102853

Khezr, P., & Mohan, V. (2022). The vexing but persistent problem of authorship misconduct in research. *Research Policy*, *51*(3), 104466. https://doi.org/10.1016/j.respol.2021.104466

Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario. (2023). *Boletín Oficial del Estado*, *70*. https://www.boe.es/eli/es/lo/2023/03/22/2/con

Martínez-Méndez, F. J., & López-Carreño, R. (2019). La paulatina adopción de ORCID para la mejora de la identidad digital de las revistas científicas españolas en acceso abierto. *Investigación bibliotecológica*, 33(80), 73-95.

Macgregor, G., Lancho-Barrantes, B. S., & Pennington, D. R. (2023). *Measuring the Concept of PID Literacy: User Perceptions and Understanding of PIDs in Support of Open Scholarly Infrastructure*. *7*(1). https://doi.org/10.1515/opis-2022-0142

Marín-Arraiza, P. (2022). Madurez de sistemas de identificadores persistentes: Oportunidades en el contexto español. *Anuario ThinkEPI*, *16*. https://doi.org/10.3145/thinkepi.2022.e16a06

Marín-Arraiza, P., & Mejias, G. (2020). Identificadores persistentes: La adopción del orcid iD en España. *Anuario ThinkEPI*, *14*. https://doi.org/10.3145/thinkepi.2020.e14e06

Meadows, A., L. Haak, L., & Brown, J. (2019). Persistent identifiers: The building blocks of the research information infrastructure. *Insights: The UKSG Journal*. https://doi.org/10.1629/uksg.457

Ministry of Science and Innovation. (2023). *National Strategy for Open Science (ENCA) 2023-2027*. https://www.ciencia.gob.es/InfoGeneralPortal/documento/e5b759a4-d756-4af9-89b0-a8cf5fd28e20

ORCID. (2024a). Retrieved December 26, 2023, *from https://orcid.org/*

ORCID. (2024b). *ORCID Statistics*. Retrieved March 25, 2024, from https://info.orcid.org/resources/orcid-statistics/

Pampel, H., Schrader, A. C., Vierkant, P., Dreyer, B., Glagla-Dietz, S., Schirrwagen, J., & Summann, F. (2024). Lessons learned from ORCID DE—A project-driven

initiative to promote author identification in Germany. *Learned Publishing*, *n/a*(n/a). https://doi.org/10.1002/leap.1597

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, *1*(1), 428–444. https://doi.org/10.1162/qss_a_00023

Porter, S. J. (2022). Measuring Research Information Citizenship Across ORCID Practice. *Frontiers in Research Metrics and Analytics*, *7*, 779097. https://doi.org/10.3389/frma.2022.779097

Porter, S. J., & McIntosh, L. D. (2024). *Identifying Fabricated Networks within Authorship-for-Sale Enterprises* (arXiv:2401.04022). arXiv. http://arxiv.org/abs/2401.04022

Quinn, A. M. (2023). Promoting ORCID registration at Emory University's School of Law. *The Journal of Academic Librarianship*, *49*(6), 102656. https://doi.org/10.1016/j.acalib.2022.102656

Schnieders, K., Mierz, S., Boccalini, S., Meyer zu Westerhausen, W., Hauschke, C., Hagemann-Wilholt, S., & Schulze, S. (2022). ORCID coverage in research institutions—Readiness for partially automated research reporting. *Frontiers in Research Metrics and Analytics*, *7*. https://www.frontiersin.org/articles/10.3389/frma.2022.1010504

Sixto-Costoya, A., Robinson-Garcia, N., Van Leeuwen, T., & Costas, R. (2021). Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: A methodological discussion. *Scientometrics*, *126*(8), 7149–7165. https://doi.org/10.1007/s11192-021-04043-5

Teixeira da Silva, J. (2020). ORCID: Issues and concerns about its use for academic purposes and research integrity. *Annals of Library and Information Studies*. https://doi.org/10.56042/alis.v67i4.36267

Teixeira da Silva, J. (2023). ORCIDs of "researchers" with the names of vegetables. *International Journal of Vegetable Science*, *29*(4), 291–293. https://doi.org/10.1080/19315260.2023.2211576

Teixeira da Silva, J. A. (2021a). Abuse of ORCID's weaknesses by authors who use paper mills. *Scientometrics*, *126*(7), 6119–6125. https://doi.org/10.1007/s11192-021-03996-x

Teixeira da Silva, J. A. (2021b). Is Mandatory ORCID a Violation of an Author's Freedoms and Rights? *Serials Review*, *47*(3–4), 243–244. https://doi.org/10.1080/00987913.2021.2022395

Teixeira da Silva, J. A. (2021c). A dangerous triangularization of conflicting values in academic publishing: ORCID, fake authors, and risks with the lack of criminalization of the creators of fake elements. *Epistēmēs Metron Logos*, *0*(7), 1–9. https://doi.org/10.12681/eml.27238

Teixeira da Silva, J. A. (2022). Should ORCID Be Mandatory for Authors, but Not for Editors? *Operations Research Forum*, *3*(1), 6. https://doi.org/10.1007/s43069-022-00125-z

Thomas, Wm. J., Chen, B., & Clement, G. (2015). Orcid identifiers: Planned and potential uses by associations, publishers, and librarians. *The Serials Librarian*, 68(1-4), 332-341. https://doi.org/10.1080/0361526X.2015.1017713

Thompson, P. T., Powell, C. D., & Moseley, H. N. B. (2022). Academic Tracker: Software for tracking and reporting publications associated with authors and grants. *PLOS ONE*, *17*(11), e0277834. https://doi.org/10.1371/journal.pone.0277834

Van Wettere, N. (2021). Affiliation Information in DataCite Dataset Metadata: A Flemish Case Study. *Data Science Journal*, *20*, 13. https://doi.org/10.5334/dsj-2021-013

Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, *60*(4), 103385. https://doi.org/10.1016/j.ipm.2023.103385

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Wang, T., Li, Z., Huang, S., & Yang, B. (2024). Is ORCID a reliable source for CV analysis? Exploring the data availability of ORCID academic profiles. *Scientometrics*, *129*(3), 1637–1662. https://doi.org/10.1007/s11192-024-04944-1

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wouters, P., Sugimoto, C. R., Larivière, V., McVeigh, M. E., Pulverer, B., de Rijcke, S., & Waltman, L. (2019). Rethinking impact factors: Better ways to judge a journal. *Nature*, *569*(7758), 621–623. https://doi.org/10.1038/d41586-019-01643-3

Youtie, J., Carley, S., Porter, A. L., & Shapira, P. (2017). Tracking researchers and their outputs: New insights from ORCIDs. *Scientometrics*, *113*(1), 437–453. https://doi.org/10.1007/s11192-017-2473-0