

ORCID iDs in the Open Knowledge Era

Marina Morgan¹[0000-0002-1829-1016] and Naomi Eichenlaub²[0000-0002-4348-1700]

¹ Florida Southern College, Lakeland, FL, 33801, USA
mmorgan@flsouthern.edu

² Ryerson University, Toronto, ON, M5B 2K3, Canada
neichenl@ryerson.ca

Abstract. The focus of this poster is to highlight the importance of sufficient metadata in ORCID records for the purpose of name disambiguation. In 2017 the authors counted ORCID iDs containing minimal information. They invoked RESTful API calls using Postman software and searched ORCID records created between 2012-2017 that did not include affiliation or organization name, Ringgold ID, and any work titles. A year later, they reproduced the same API calls and compared with the results achieved the year before. The results reveal that a high number of records are still minimal or orphan, thus making the name disambiguation process difficult. The authors recognize the benefit of a unique identifier that facilitates name disambiguation and remain confident that with continued work in the areas of system interoperability and technical integration, alongside continued advocacy and outreach, ORCID will grow and develop not only in number of iDs but also in metadata robustness.

Keywords: ORCID, Open Data, Public Data, Persistent Identifiers, Researcher Identifiers, Interoperability, Name Disambiguation, Metadata Assessment, API.

1 ORCID, ORCID iDs, and Name Disambiguation

1.1 ORCID and ORCID iDs

ORCID is a non-profit organization that has facilitated an open registry of persistent digital identifiers since 2012. ORCID is backed by a vast community comprised of research organizations and institutions, publishers, professional associations, individual research and organizational members.

ORCID iD, the Open Researcher and Contributor Identifier, is a persistent digital identifier provided by ORCID to researchers across disciplines. The purpose of the ORCID iD is to distinguish researchers by accrediting their research and activities, and improving recognition and discoverability. The only required metadata fields for creating and registering an ORCID iD are first name and primary email address [1]. The researcher has the flexibility of controlling what data is entered in their ORCID iDs, which can be made private, public, or shared with a limited group. It is the researcher's responsibility to create and maintain their ORCID record as ORCID is not in a position to modify any incorrect data [2].

1.2 Name Disambiguation

Author name remains an ongoing challenge in the field of scholarly communication. Extensive changes to authorship and collaboration in recent decades have resulted in an acceleration in multi-authorship. These changes can be attributed in large part to the present day facility for online collaboration but also to an increase in interdisciplinary collaboration and a heightened focus on research and publication outputs. As such, today it is not uncommon for science articles to be authored by hundreds of authors [3], but the trend of increasing co-authorship applies in the social sciences and humanities as well [4, 5]. A further challenge has been the growth in research output from countries where authors have recurrent similar names [6]. All of these authorship issues impact researcher identity and increase the urgency for author name disambiguation.

2 Methodology and Results

2.1 Background

Approaching from the perspective of exploring the quality and utility of author name metadata, in 2017 the authors investigated and queried the ORCID records and metadata fields using the public ORCID API to evaluate the completeness of metadata in ORCID records [7]. ORCID iDs are useful in the author disambiguation process only if they provide enough information to distinguish one author from another. Thus, the focus of the investigation was to count the records with only the minimal required information, specifically name and email. The search focused on records created from 2012 to May 2017 that did not include affiliation name, any work titles, and Ringgold ID (a numerical identifier assigned to scholarly institutions). The results revealed that approximately 65% of the ORCID records were minimal, or “orphan”. Moreover, it was observed that some records included false or misleading metadata such as random names (John Doe or Jane Doe), misleading funding (Awards-R-U or Grant-R-U), and filler text in the biographies (lorem ipsum), further hampering disambiguation.

2.2 Methodology and Results

To get a better sense of how many ORCID records have only minimal information and in order to generate the results needed for the comparison with the original results, we reran the same public ORCID API calls for records created between 2012 and 2017. The 2017 count includes records created between January and May. The same authentication token and Postman software were used to invoke RESTful API calls.

As seen in Table 1, even though there is a slight improvement in the number of records that do not include affiliation name, Ringgold ID, and any work titles, orphan or empty records are still prevalent. Additionally, the results reveal that on average

there has been a 6.5% improvement, an unexpectedly low percentage. However, this does not come as a surprise as ORCID “does not absolutely prevent multiple iDs/records from being intentionally created and maintained by an individual” [8]. Despite ORCID’s policies in place to avoid duplicate records, the high number of orphan records may be because of duplicates since individuals who have multiple email addresses can in theory create as many records as email addresses.

Table 1. Comparison of Public API calls for minimal ORCID records for 2017 and 2018.

Year	2017 Count	2018 Count	Improvement Percentage
2012	25,351	23,724	6.42%
2013	258,182	239,582	7.20%
2014	370,074	344,213	6.99%
2015	479,144	448,453	6.41%
2016	709,046	666,447	6.01%
2017*	372,709	344,405	7.59%
2012-2017*	2,216,944	2,070,491	6.61%

* January – May 2017

Table 2 illustrates a total count of ORCID records starting with 2012. The 2017 count revealed that approximately 65% of the total records were minimal. When the count was repeated in April 2018, the results showed that nearly 45% are minimal. This decrease from 65% to 45% represents a significant reduction. However, with just under half of ORCID records remaining without enough metadata for proper disambiguation, there is still room for improvement.

Table 2. Total ORCID Records

Date	Total ORCID Records	Minimal ORCID Records	Percentage of Minimal ORCID Records
2012 - May 17, 2017	3,391,358	2,216,944	65.37%
2012 - April 03, 2018	4,625,545	2,070,491	44.76%

The total number of ORCID records continues to grow at an impressive rate of 36.4% from May 2017 to April 2018. Therefore, the rapid uptake of ORCID by researchers shows that while many records do not have adequate metadata, there is the potential for both comprehensive researcher uptake as well as more robust and complete associated metadata in the future [9]. ORCID initiatives such as Collect & Connect, which aim to validate research affiliations through authentication, have likely contributed to the reduction in orphan records shown in Table 2 [10].

3 Conclusions

This project involved a follow-up investigation of work done in 2017 to determine the completeness of metadata in ORCID records. Specifically, the authors used API calls to conduct searches to count records created between 2012 and May 2017 that did not include affiliation name, any work titles, and Ringgold ID. API calls were repeated in April 2018 in order to determine changes in counts. The 2017 count revealed that more than 65% of the total records were minimal. When the count was repeated in April 2018, the results showed that approximately 45% are minimal, a substantial decrease.

Since ORCID does not modify incorrect data, but they may correct invalid data such as empty or wrongly formatted fields, this improvement is likely due to the researchers' recognition of the importance of accurate data, or perhaps adding additional metadata to a record that was previously created.

Lastly, the authors recognize the benefit of a unique identifier that facilitates name disambiguation and the ongoing ORCID initiatives to increase metadata robustness and record quality through initiatives such as Collect & Connect.

References

1. Building your ORCID record & connecting your iD <https://support.orcid.org/knowledgebase/articles/142948-name>, last accessed 2018/06/18.
2. ORCID Trust, <https://orcid.org/about/trust/integrity>, last accessed 2018/06/18.
3. King, C.: Multiauthor papers: onward and upward. ScienceWatch Newsletter (2012). http://archive.sciencewatch.com/newsletter/2012/201207/multiauthor_papers/, last accessed 2018/04/05.
4. Taylor & Francis Group, Co-authorship in the Humanities and Social Sciences: A global view. A white paper from Taylor & Francis, (2017). <http://authorservices.taylorandfrancis.com/wp-content/uploads/2017/09/Coauthorship-white-paper.pdf>, last accessed 2018/04/05.
5. Kuld, L., O'Hagan, J. The proportion of co-authored research articles has risen markedly in recent decades. LSE Impact Blog, (2018). <http://blogs.lse.ac.uk/impactofsocialsciences/2018/04/04/the-proportion-of-co-authored-research-articles-has-risen-markedly-in-recent-decades/>, last accessed 2018/04/05.
6. Youtie, J., Carley, S., Porter, A.L., Shapira, P.: Tracking researchers and their outputs: new insights from ORCIDs. *Scientometrics* (113), 437 (2017).
7. Eichenlaub, N., Morgan, M.: ORCID: Using API Calls to Assess Metadata Completeness. In: Proceedings of the International Conference on Dublin Core and Metadata Applications, pp. 104-107. Dublin Core Metadata Initiative, Washington D.C. (2017).
8. Managing Duplicate ORCID iDs, <https://orcid.org/blog/2014/01/09/managing-duplicate-iDs>, last accessed 2018/04/05.
9. Klein, M., Van de Sompel, H.: Discovering scholarly orphans using ORCID. In: 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). (2017).
10. ORCID. Welcome to Collect & Connect: ORCID's integration and engagement program, <https://orcid.org/content/collect-connect>, last accessed 2018/04/10.