ORCID: Using API Calls to Assess Metadata Completeness
Poster

Naomi Eichenlaub
Ryerson University, Canada
neichenl@ryerson.ca

Marina Morgan
Florida Southern College, United States
mmorgan@flsouthern.edu

Keywords: ORCID; persistent identifiers; researcher identifiers; interoperability; APIs; public data; name disambiguation; metadata assessment.

Abstract

The aim of this poster is to demonstrate the importance of adequate metadata in ORCID profiles to ensure name disambiguation. It is only through more complete metadata that ORCID will ensure success in terms of interoperability with institutional scholarly, publishing and funding bodies.

Introduction

Launched in 2012, ORCID (Open Researcher and Contributor ID) is a non-profit persistent digital identifier open registry offered to researchers across disciplines. Their mission is to provide an identifier in the form of a unique alphanumeric code to provide persistent identity for researchers. There are three easy steps to receive a persistent ID and distinguish oneself from other researchers: register, add your info, and use your ORCID ID. However, we noticed that a very large number of ORCID IDs are empty, i.e., a name is registered and a profile is created but they lack critical elements required to perform the function of name disambiguation such as country and affiliated institution. Other metadata, for instance activities summary, funding, peer-reviews, and works can be considered important but not critical to the primary function of ORCID, which is name disambiguation. ORCID itself has noted this problem in a blog post in early 2017 stating that "Many records have limited (public or trusted party) information beyond the author's name and iD" (ORCID, 2017a). Duplicate records (created when an author forgets they have already created an ORCID) are a much smaller issue and are being addressed by ORCID with new initiatives such as self-management of duplicate records (ORCID, 2017a). Essentially, the richer the metadata the more useful something becomes so it behooves the research community as a whole to work together to make ORCID records as rich and complete as possible. Therefore, in an effort to get a better sense of the overall completeness of the current state of ORCID records, we decided to investigate and query the records and metadata fields using the public ORCID API.

Methodology and Results

We ran queries against the public ORCID API to get a better sense of how many ORCID records have only minimal information. In order to query the public file, we registered for a public API client application, enabled the developer tools for the application, and accessed the authorized URL to retrieve an authentication code. Once we accessed the token URL, we retrieved an authentication token, which was used for all the API calls. To invoke RESTful API calls we used Postman software that enabled the following headers: Accept: application/json, Authorization_type: Bearer, and the Access_token.

To get a sense of the big picture in terms of records with only the minimal required information, i.e., name and email, we searched records created between 2012-2017 that did not

include affiliation / organization name, Ringgold ID, and any work titles. Results were gathered by year as well as an overall API call to search the entire ORCID database.

TABLE 1: Examples of public API calls.

Year	API Calls	Count
2012	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st to December 31st, 2012: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2012-01-01T00:00:00Z%20TO%202012-12-31T00:00:00Z%5D	25,351
2013	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st to December 31st, 2013: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2013-01-01T00:00:00Z%20TO%202013-12-31T00:00:00Z%5D	258,182
2014	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st to December 31st, 2014: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO*]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2014-01-01T00:00:00Z%20TO%202014-12-31T00:00:00Z%5D	370,074
2015	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st to December 31st, 2015: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2015-01-01T00:00:00Z%20TO%202015-12-31T00:00:00Z%5D	479,144
2016	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st to December 31st, 2016: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2016-01-01T00:00:00Z%20TO%202016-12-31T00:00:00Z%5D	709,046
2017 to 05/17/17	ORCID records with any given names, without any affiliation, Ringgold ID, work titles, and submitted between January 1st, 2017 to May 17, 2017: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]+AND+profile-submission-date:%5B 2017-01-01T00:00:00Z%20TO%202017-12-31T00:00:00Z%5D	372,709
2012 to 05/17/17	All ORCID records with any given names, without any affiliation, Ringgold ID, and work titles: https://pub.orcid.org/v2.0/search/?q=given-names:[* TO *]+AND+-affiliation-org-name:[* TO *]+AND+-ringgold-org-id:[* TO *]+AND+-work-titles:[* TO *]	2,216,944

The API calls were made on May 17, 2017. The results as seen in the above table are relevant to that date and may have changed since then. Current ORCID statistics can be viewed at the following URL: https://orcid.org/statistics.

Moreover, based on the above results and the total number of ORCID records submitted between 2012 and 2017 (up to May 17, 2017) we calculated the percentage of minimal ORCID records in the respective year. More than half of the records submitted are minimal.

TABLE 2: Percentages of minimal ORCID records by year.

Year	Total ORCID Submissions	Minimal ORCID Records Count	Percentage
2012	44,118	25,351	57.46%
2013	424,927	258,182	60.75%
2014	608,999	370,074	60.76%
2015	769,979	479,144	62.22%
2016	1,049,820	709,046	67.53%
2017*	948,445	372,709	39.29%

* Up to May 17, 2017

Challenges

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2017

Missing data such as affiliations or work titles poses a major challenge on one hand to gather appropriate data and on the other to consider PID service adoption. From the preliminary results we noticed that the problem of orphan (empty) records is very common. Many records used random placeholder names such as John/Jane Doe to keep their true identity unknown, filler text such as lorem ipsum, fictitious funding, works, and institution names for employment and education.

At this stage in our research it is beyond our purpose to search how many names are similar and/or unable to be disambiguated due to lack of additional information even though technically not impossible.

In a blog post on the challenges of measuring PID (Persistent IDentifier) adoption by Robin Dasler (senior fellow in CERN's Scientific Information Service) she pointed out that ORCID acknowledges that when the service first launched, "it was fine to be concerned only with uptake, since the priority was to get the word out." However, with the growth and development that has occurred over the years, the focus needs to be on attaining innovation and interoperability (ORCID, 2017b).

In an effort to foster integration and engagement within the research community, ORCID launched the Collect & Connect program in 2016. With increased trust in connections "between researchers and their professional affiliations and activities" a greater number of ORCID identifiers can be collected and connected (Meadows, A., 2016), thereby maximizing metadata robustness and interoperability (ORCID, 2017c).

Conclusions and Future Work

ORCID IDs are very useful, specifically when most names are not unique. However, many records lack critical elements required to perform the main function of a personal identifier, namely the name disambiguation.

To this end, the authors propose that one of the priorities going forward should be to work together to ensure that a greater number of ORCID records have a higher number of completed metadata fields especially since incomplete metadata poses a challenge to name disambiguation. This issue has been acknowledged by ORCID as well: "We need to do more to ensure that ORCID identifiers are collected using appropriate, validated methods, and are published with research activities and affiliations" (Meadows, A., 2016).
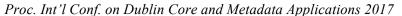
In order to address a large number of ORCID records missing critical metadata fields the following solutions are proposed:

1. Advocacy and education - via their unique positions, librarians in academic settings can provide research support services such as individual and personalized researcher profile consultation services offered to targeted researchers at an institution in order to ensure ORCID records are more complete (Thompson, E. & French, S., 2017; Reed, K., McFarland, D., & Croft, R., 2016), and can lead "campus-wide efforts to promote the use of ORCID and similar resources" (Tran, C.Y. & Lyon, J.A., 2017).

2. Continued ORCID outreach: ORCID has identified a "key goal" for 2017 "to develop education and outreach resources for researchers explaining how and why to connect information in the ORCID records" (ORCID, 2017a).

3. Increased system interoperability: ORCID notes that they are "working with third party system vendors to improve information flow between systems" and that they will "continue to expand the types of works and activities that can be connected to ORCID records" (ORCID, 2017b).

It is hoped that this poster draws attention to the problems associated with a lack of identifying metadata in ORCID records and highlights the value of more complete metadata in disambiguating researcher names and identifiers.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2017

References

Meadows, Alice. (2016). Everything you ever wanted to know about ORCID … but were afraid to ask. College & Research Library News, 77(1), 23-30.

ORCID. (2017a). ORCID open letter: One year on. Retrieved, May 15, 2017, from https://figshare.com/articles/ORCID_Open_Letter_-_One_Year_On_Report/4828312.

ORCID. (2017b). Challenges of measuring PID adoption. Retrieved, April 17, 2017, from http://orcid.org/blog/2017/03/24/challenges-measuring-pid-adoption.

ORCID. (2017c). Collect & Connect. Retrieved, March 15, 2017, from https://orcid.org/content/collect-connect.

ORCID. (2017d). ORCID API. Retrieved, March 15, 2017, from https://orcid.org/organizations/integrators/API.

Reed, Kathleen, Dana McFarland, and Rosie Croft. (2016). Laying the groundwork for a new library service: Scholar-practitioner & graduate student attitudes toward altmetrics and curation of online profiles. Evidence Based Library and Information Practice, 11(2), 87-96. doi:10.18438/B8J047.

Ringgold. (2017). Case Studies ORCID. Retrieved, March 15, 2017, from http://www.ringgold.com/case-studies-orcid.

Thompson, Elleb and Sally French. (2016). Pimp my Profile and the researcher profile health check: Practical, individualised researcher support initiatives co-created by library and faculty. In ALIA National Conference 2016, 29 August - 2 September 2016, Adelaide, SA. Retrieved, March 15, 2017, from http://eprints.qut.edu.au/98649/.

THOR. (2017). Researcher metrics: ORCID IDs. Retrieved, March 15, 2017, from http://dashboard.project-thor.eu/dashboard/researcher/.

Tran, ClaraY. and Jennifer A. Lyon. (2017). Faculty use of author identifiers and researcher networking tools. College & Research Libraries, 78(2), 171-182. doi:10.5860/crl.78.2.171.