ORCID FOR **FUNDING ORGANIZATIONS**

GRANT APPLICATION REPORTING

ORCID | ORBIT: Grant Application Data Field Survey Report

JULY 2019

PREPARED BY:

JOSH BROWN (ORCID)
ⓘD https://orcid.org/0000-0002-8689-4935

LAUREL L. HAAK (ORCID)
ⓘD https://orcid.org/0000-0001-5109-3700

TOM DEMERANVILLE (ORCID)
ⓘD https://orcid.org/0000-0003-0902-4386

ALICE MEADOWS (ORCID)
ⓘD https://orcid.org/0000-0003-2161-3781

TABLE OF CONTENTS

1.0 REPORT BACKGROUND AND PURPOSE

This report summarizes the results of the ORBIT Working Group survey of grant application data needs carried out in 2017, in which nine research funding bodies from four continents participated.[1] These included large national multidisciplinary funders from seven countries and discipline-focused philanthropic funders from two countries. The report incorporates comments from the group and has been endorsed for publication as an output of the group.

2.0 EXECUTIVE SUMMARY

This report summarizes the findings of a survey of data fields used by funders to collect grant application information. It was carried out in 2017 as part of the ORBIT project, and represents responses from nine participating funders.

Based on their responses, in this report we compare data fields currently supported within the ORCID record with the lists of data fields that funders collect during the grant application process. We also analyze the limitations and opportunities presented by the current research information ecosystem and highlight challenges in reusing that data, as well as flagging types of information for which additional ORCID fields may be required.

Our analysis shows that the ORCID data model accommodates — or could easily accommodate — much of the grant applicant information required by funders. The analysis has led to specific actions for both ORCID and funders, and its conclusions may be more broadly useful in highlighting actions to maximize the availability of open, reusable funding information, in particular through the use of open persistent identifiers (PIDs) and metadata.

Ultimately, our goal is to enable researchers to easily share information about their activities and affiliations with grant application systems, reducing the data entry burden for them and improving data quality for funders and the broader community.

---

[1] Initial version completed 26 Feb 2018, Authors: Josh Brown and Tom Demeranville.
The current version incorporates comments and analysis by Laure Haak and Alice Meadows.

 ORCID

3.0 METHODS

The first step in the survey analysis was to take the mappings of fields collected by funders and held in the ORCID Registry, and subdivide them based on the component sections of an ORCID record. The 'gold standard' for ORCID is data from an authoritative source (e.g., employment affiliation asserted by the employee's organization, or works information asserted by the publisher) including a persistent identifier. Where such data are available reliably (currently or potentially) and are already supported by the Registry, our analysis ends. Where there are challenges specific to an information type (such as a lack of integrated sources, or a lack of open information and/or identifiers), the scale and nature of these challenges are assessed.

The second step in the analysis was to review the current data available in the Registry, and its provenance. We used these data to generate a preliminary analysis of the challenges and likely timescale of extending the coverage and availability of data via the Registry or via other open community resources.

Our analysis focuses on fields that are currently supported by the ORCID data model.[2] This document uses the following criteria to categorize information:

- **Supported:** Data field used by funder system is currently supported by the ORCID data model and existing ORCID integrations by community information platforms

- **Integration needed:** Data field used by funder system is currently supported by the ORCID data model but requires new integrations with existing authoritative sources

- **Source needed:** Data field used by funder system is currently supported by the ORCID data model but authoritative sources are not available

- **Privacy conflict:** Data field used by funder system conflicts with ORCID privacy principles and will not be supported by the ORCID Registry

---

[2] Note that the initial survey was completed in 2018; since then ORCID has released API3.0 which includes new data fields that address some of the requirements identified in this survey. These changes are addressed in-line in the text.

 ORCID

4.0 IDENTIFYING PRIORITIES

This section reviews categories of information fields collected in grant application systems, with the goal of identifying data fields and prioritizing information sources that ORCID can partner with to streamline data re-use for researchers. We highlight those fields that are already well-supported by the Registry, and flag those fields that are an obvious priority for addition to the ORCID data model, namely those which a majority of funders collect but that ORCID does not.

This analysis serves to eliminate three kinds of data fields from the prioritization process:

   (i) fields that are already supported in ORCID data model

   (ii) fields that are only useful to a small subset of funders

   (iii) fields that will not be supported by the ORCID Registry in the foreseeable future

Note that the final category is revisited under the source analysis and discussions in sections 5 and 6.

There will, of course, be amendments and additions to this analysis, such as a field that most funders may not currently collect but which they would regard as high value should it be reliably supported by ORCID and others.

4.1 BIOGRAPHICAL INFORMATION

NAME INFORMATION

The ORCID data model accommodates given names, family names, a credit name, and other associated names. These fields support the 'first name' and 'last name' grant application information requirements of all of the respondents (n=9). In addition, three additional fields were indicated by some of the respondents: 'title' (n=4), 'middle name,' (n=4) and 'salutation' (n=2). These are all good candidates for inclusion in the Registry. Note that the ORCID credit name field is where middle names or initials are currently recorded, but this is part of the name as a whole, not a separate field.

Apart from 'other names,' name information is added manually by the ORCID record owner. If ORCID were to update our 'name' data model, therefore, a communications effort by funders would be needed to encourage or require researchers to make these changes.

PERSON IDENTIFIER INFORMATION

The ORCID data model includes other person identifiers. The intent is for these to be (i) directly relevant to the record holder's research career, and (ii) appropriate for sharing in the public domain. Survey respondents mentioned collecting ORCID iDs (n=5) and also Scopus Author ID (n=2), Researcher ID (n=1), eRA Commons ID (n=1), and the Japan Researcher ID (n=1). All of these are public person identifiers and are supported in the existing ORCID data model. Some respondents mentioned a Healthcare Provider ID and Canadian social security numbers, which are private identifiers and as such conflict with ORCID's privacy policy.

ORCID

ORCID allows members to connect resolvable person identifiers with ORCID records (with the permission of the record holder). In practice, the majority of person identifiers connected to ORCID records are Scopus IDs or Researcher IDs. While record owners cannot add person IDs directly, they can initiate this process, for example, via the Scopus or Publons (formerly ResearcherID) search and link wizards.

The ORCID data model also accommodates researcher website URLs: profile systems, personal blogs, research group pages, twitter handles, and the like. This information can be added by both members and record-holders.

LANGUAGE INFORMATION

Language preferences were noted by two respondents. This is not a Registry data field that ORCID currently supports. ORCID does provide a multi-language user interface, and we capture browser language preferences for displaying records to users in their language of choice (where available).

'PERSONAL' INFORMATION

Some respondents require information of a personal nature in grant application workflows, such as 'place of birth' (n=1), 'date of birth' (n=4), age (n=2), 'gender' (n=4), 'ethnicity (n=2)', 'disability' (n=1), 'citizenship/nationality' (n=2), and 'address' (n=5).

Collecting or storing these data fields conflicts with ORCID privacy principles. They present a higher level of risk and consequent regulation in the global data protection environment, and therefore pose practical and legal challenges specific to ORCID as a global source of data. Furthermore, race and ethnicity fields are relevant only in a local context, while gender identity and disability have field values that are not globally standardized.

ORCID operates on open principles. Our Registry is a hub for public data. Data that cannot be made open or public appropriately or safely via the Registry should not be shared in ORCID records. The desire to reduce researcher burden should not run counter to the need for individuals to evaluate, control, and manage requests for sensitive personal information.

Given that ORCID cannot realistically be a hub for all the possible information the research community might ever require (and nor should it seek to be), we see a role for secure local systems (such as a university human resources system) to complement ORCID Registry data, but behind their firewall.

OTHER INFORMATION

Keywords. Most respondents collect keywords of some kind (n=5), often centered on research area/discipline/focus, and sometimes taken from a fixed vocabulary. ORCID currently supports a generic free text 'keywords' field. Lacking a globally relevant ontology for research fields, the record-holder or their institution can add a shared keyword, research area, or community name to their keywords section. ORCID is investigating options for additional work in this area to support regional ontologies.

ORCID

Status. Some form of status field was required by two respondents. Active, lapsed, and career status were all mentioned. This information is very specific to individual system needs and definitions (for example, there is no global definition of early career researcher). While status fields are not specifically supported in the ORCID data model, existing start and end dates for affiliations could be utilized by grant application systems to supplement these fields.

Biographies. Interest in narrative biographies was expressed during ORBIT meetings, and four respondents indicated a requirement for some form of personal statement, which could be seen as a specialized form of biography. A biography field is supported in the ORCID data model, however, it may not be suitable in the current form for the needs of grant application systems as this field is usually very specific to an application or a current moment in time.

4.2 EDUCATION

The ORCID data model includes many of the education fields required by respondents, including institution, degree, start/end dates, degree/title, thesis title. However, components of an education activity are aggregated by funding systems in different ways. For example, the ORCID data model accommodates start and end dates for educational activities. It is unclear how this maps to funder system requirements for 'course year' or 'expected date.'

Some funding systems store course title separately from the grade, while ORCID stores much of this information in a compound 'title' field. In addition, notes about grades cannot be stored in the ORCID Registry. These are probably best made on a case-by-case basis in the educational institution's system. This means that some kind of data model translation is required by the funder system to enable researchers to re-use data from their ORCID record.

The ORCID data model specifies thesis/dissertation as a 'work', distinct from 'education.' Since a thesis/dissertation is a scholarly contribution, it does make sense to specify it as a 'work' rather than an 'affiliations/education' item in our data model. We have seen some universities start to assert information into ORCID records when a researcher graduates. As more institutions follow this practice, researchers will be able to easily share education information with funding systems. Specifically, institutions can:

(i) assert educational degree, date, and awarding institution into the affiliation section of the researcher's ORCID record

(ii) assert thesis/dissertation, publication date, and associated DOI in the works section of the researcher's ORCID record

(iii) acknowledge service, with a link to the thesis/dissertation DOI in the ORCID record for each of the members of the thesis/dissertation committee

ORCiD

4.3 EMPLOYMENT

The ORCID data model supports most of the information about employment activities required by respondents, including institution/employer, school/department, location, role, start/end dates. As noted above, start/end dates can help funding systems discern status. In addition, ORCID uses organization identifiers, which can enable systems to discern location details and organization type. Rank/tenure information was indicated by two respondents; this information may be recorded as free text but, given varying geographical practices, this is likely to be of limited utility and may be better as a field for the funder system to collect.

4.4 FUNDING

The ORCID data model accommodates the majority of funder system requirements, including funder name, funder location, grant title, start/end dates, amount. The only omission mentioned by more than one funder is 'percent effort', or the proportion of a grantee's time devoted to working on the research proposed in the grant. There is a trend toward the use of persistent identifiers by funders for their grant awards, which will enable streamlined and high-fidelity information sharing.[3]

4.5 PEER REVIEW

Only one respondent indicated an interest in collecting peer review activity in a grant application workflow. The ORCID data model includes peer review service, which is currently used by publishers and third party services, and is also available for use by funders and other organizations that undertake forms of peer review.

4.6 OTHER ACTIVITIES

Respondents also mentioned other professional activities, including professional organization membership, certifications, distinctions, awards, mentorship/supervision, and other contributions. At the time of this survey, ORCID was working with the community to develop a structured data model for these types of activities (Appendix D). Input from this survey was incorporated, and an expanded affiliations data model was launched with API version 3.0 in 2019, which enables the addition of information about non-academic qualifications, membership, service, invited positions, and distinctions.

Other activities mentioned included licences and leaves of absence. Of these, licences are supported in the ORCID data model as a work activity. ORCID considers information on leaves of absence to be in conflict with our privacy policy.

---

[3] See, for example, the ORBIT Working Group recommendation on ORCID and Grant DOI interactions.

ORCID

5.0 SOURCE ANALYSIS

While the ORCID data model may accommodate information required in funding application systems, it is important that, for the purposes of reduction of researcher burden and data quality/fidelity, these data are added to ORCID records as a researcher interacts with various information systems. In this section, we review the amount and sources of information in the ORCID Registry and identify priorities for targeted ORCID integrations.

We used the 2017 and 2018 ORCID public data files to analyze the source of Registry items, to establish whether data were API-asserted (that is, added during researcher-initiated system-to-system interactions) or self-asserted (that is, added directly by researchers). The results are summarized in Tables 1 and 2 below. While the number of assertions increased in 2018 for each field type, the proportion of assertions made through the API vs. self-asserted remained fairly stable across the two years. Note that peer review activities and person IDs cannot be self-asserted, they can only be added via a member API integration.

TABLE 1. ORCID REGISTRY ITEM VOLUME AND COUNT, OCTOBER 2017

OCT 2017	EDUCATION	EMPLOYMENT	FUNDING	PEER REVIEW	WORK	PERSON ID
Total count	1,672,400	1,370,284	285,908	128,432	21,789,366	744,012
Total added via member API	10,214	63,439	95,995	128,431	18,455,734	744,011
% added via member API	0.6%	4.6%	33.6%	100.0%	84.7%	100.0%

TABLE 2. ORCID REGISTRY ITEM VOLUME AND COUNT, OCTOBER 2018.

OCT 2018	EDUCATION	EMPLOYMENT	FUNDING	PEER REVIEW	WORK	PERSON ID
Total count	2,551,324	2,038,682	443,274	459,110	31,132,068	1,037,202
Total added via member API	14,895	68,063	142,729	459,109	26,184,074	1,037,201
% added via member API	0.6%	3.3%	32.2%	100.0%	84.1%	100.0%

ORCiD

5.1 EDUCATION AND EMPLOYMENT

The vast majority of education and employment assertions are added through the user interface (self-asserted). Of the four million or so ORCID records in the 2017 data file, 28% included an education, employment, or both affiliation types. In 2018, there were over five million records, and 30% included an education, employment, or both affiliation types. These affiliations are more common than any other data type. Around 99% of the education assertions and 95% of the employment assertions were made by the ORCID record-holder using the web user interface. Their institutions are just starting to assert these employment and education affiliations to ORCID, as are third party 'faculty profile' systems. During 2017, we focused our efforts on encouraging institutions to assert affiliations. The number of institutions doing so increased from 72 institutions adding employment data in the 2016 datafile to 127 in 2017, and the number of institutions adding education data increased from 12 in 2016 to 42 in 2017. This trend has continued and we anticipate seeing proportionately more API assertions for education and employment in 2019 and later.

5.2 FUNDING

Nearly all funding metadata is self-asserted by researchers. Of the ORCID records in the 2017 data file, 2.2% included at least one funding activity; in 2018 this had increased to 2.5%. Across years, nearly two-thirds of funding metadata were entered manually by researchers. Most of the remaining third was asserted into ORCID records via user-initiated imports using a 'search and link wizard' provided by Uberwizard. The Autism Speaks Grants system deserves a special mention as the first funder system to connect directly to ORCID and actively add grant information to records. Otherwise, funders were absent from the list of sources populating ORCID records with funding information in 2017-18. This is one of the core challenges being addressed by the ORBIT project. In 2019, more funders started to assert grant information into ORCID and, with more coming on line in 2020, we anticipate the proportion of API-asserted funding information to grow.

5.3 PEER REVIEW

Peer review is a growing activity type in the ORCID record. Of the ORCID records in the 2017 data file, 0.2% included least one peer review activity; this doubled in 2018. ORCID worked closely with the research community to define a peer review specification, which requires that any peer review information be added by external sources. By far the main source of peer review information in ORCID records has been Publons, a third-party publishing peer review recognition service. In addition, a growing group of publishers is asserting peer review information to ORCID records, most notably F1000 and the American Geophysical Union's Geophysical Electronic Manuscript Submission 'GEMS,' which have each asserted thousands of peer review activities. ORCID is actively engaging with manuscript submission systems, publishers, and funders to encourage wider adoption, so we are hopeful that there will be more use of this functionality in future.

ORCID

5.4 WORKS

Most work activities are asserted using the ORCID API. Of the ORCID records in the 2017 data file, 19.5% included at least one work; this increased to 21.9% in 2018. Journal articles, conference contributions, book contributions, and theses/dissertations are all well represented within the Registry. Datasets, patents, and over 20 other work types are supported by ORCID, but less well-used. Over 84% of work metadata are asserted into ORCID records using the ORCID API, from a diverse range of sources. There are over 150 organizations contributing work metadata to ORCID. Of these, the most prolific contributors by far are the large indexing and metadata services, largely via user-initiated imports using the 'search and link wizard' model. These include Scopus, ResearcherID, and Crossref, as well as the discipline-specific services provided by NASA Astrophysics, Europe PubMed Central, and Inspire High Energy Physics. In addition to these services, many specialist and institutional ORCID members have connected their repositories with ORCID. We recently partnered with the repositories community to define repository best practice recommendations and are actively nurturing repository integrations to increase the proportion of datasets linked from the Registry. Europe PubMed Central, along with The Lens and around 40 institutions are all enabling researchers to connect their patent activities to their ORCID records.

5.5 PERSON IDENTIFIERS

Across years, about 15% of ORCID records are associated with at least one person identifier, usually Scopus ID, Researcher ID, or both. ORCID record-holders cannot add person ID information themselves, and must rely on information-sharing by external sources. Record-holders can and do add other person information. Other names are generally added by researchers themselves (<80%), with the notable exception being services and institutions based in China, Hong Kong, and Taiwan, which make up the entire top five sources. Keywords are almost always added by researchers, (<95%), although some integrations are starting to do so; in 2017, 50 members had added at least one keyword. Nearly 100 members were asserting website links to the ORCID Registry in 2017, accounting for 20% of website metadata. The majority of these were research institutions, likely linking to institutional profile pages. The largest contributor, and also the only non-institutional contributor, is Mendeley with almost 50,000 assertions to Mendeley profile pages.

5.6 NEW SOURCES

With the launch of the ORCID API 3.0 in 2019, we are starting to see the addition of affiliation and research resource information. This includes fields requested by funding systems such as society memberships and certifications, as well as information on the use of research resources, many of which are major capital investments for funders and very difficult if not impossible to track. The ORCID research resources specification resulted from a consultation with facilities and other resource providers in collaboration with publishers.

ORCID

6.0 SUMMARY AND NEXT STEPS

Our analysis shows that the ORCID data model accommodates — or could easily accommodate — much of the grant applicant information required by funders. At present, applicants need to enter this information manually into funder systems, which is needlessly time-consuming.

ORCID can help by engaging with funders and the system providers they use to enable researchers to share information between ORCID and the application systems. This entails developing and maintaining technical documentation, providing venues for testing and promoting effective practices, and ongoing efforts to support open and transparent processes for sharing research information.

Funders can help by considering ORCID as a data source and also by more openly sharing funded award system and applying best practices for use of persistent identifiers for grantees and grants.

To this end, the ORBIT project continues to engage with the funding community to develop a shared understanding and work to develop an open research information infrastructure. Please see the subsequent survey of reporting systems data needs, and the ORBIT Working Group recommendation on ORCID and Grant DOI interactions.

See the ORBIT page on our website for more information implementing ORCID, see: https://orcid.org/organizations/funders/orbit.

ORCID

7.0 APPENDIX: ORCID FIELD USE

TAKEN FROM THE 2017 ORCID PUBLIC DATA FILE (3920204 TOTAL RECORDS):

Field / Field Type	Number of records	% of records with at least one item of field type
Employment	818577	20.9%
Education	941386	24.0%
Work	764445	19.5%
Funding	86925	2.2%
Peer review	9445	0.2%
Any affiliation	1112410	28.4%
Any activity	1396143	35.6%
Any person identifier	581728	14.8%
Affiliation and work	482515	12.3%
Funding and work	70621	1.8%
Funding and affiliation	83245	2.1%
Funding and affiliation and work	67572	1.7%
Public email	40146	1.0%
Country	575760	14.7%
Researcher ID	348065	8.9%
Scopus ID	331127	8.4%
Given name	3909379	99.6%
Family name	3842796	98.0%
Credit name	300045	7.6%
Other name	239001	6.1%
Keyword	278544	7.1%
Researcher URL (website link)	296030	7.5%

ORCID

TAKEN FROM THE 2018 ORCID PUBLIC DATA FILE (5292284 TOTAL RECORDS):

Field / Field Type	Number of records	% of records with at least one item of field type
Employment	1183742	22.4
Education	1410047	26.6%
Work	1159441	21.9%
Funding	134702	2.5%
Peer review	22258	0.4%
Any affiliation	1627843	30.7%
Any activity	2031273	38.4%
Any person identifier	792157	15%
Affiliation and work	759008	14.3%
Funding and work	111623	2.1%
Funding and affiliation	129729	2.5%
Funding and affiliation and work	107474	2%
Public email	55186	1%
Country	790673	15%
Researcher ID	452537	8.6%
Scopus ID	481230	9.1%
Given name	5275282	99.7%
Family name	5187358	98%
Credit name	202215	3.8%
Other name	305216	5.8%
Keyword	372824	7%
Researcher URL (website link)	422572	8%

ORCiD