Using ORCID to track awardees’ progress

Gary McDowell, PhD
10/06/2022

Financial COIs:
I work as a freelance academic consultant who is paid for this line of work on a per-client basis.
I have no current or prior financial (or otherwise) connection with ORCID.
What is ORCID?

The ORCID iD: a unique, persistent identifier free of charge to researchers

An ORCID record connected to the ORCID iD, and

A set of Application Programming Interfaces (APIs), as well as the services and support of communities of practice that enable interoperability between an ORCID record and member organizations so researchers can choose to allow connection of their iD with their affiliations and contributions

See: https://info.orcid.org/what-is-orcid/
Why ORCID is suitable for outcomes tracking

- ORCID is useful because:
  - it’s a **unique**, permanent ID for researchers
Why ORCID is suitable for outcomes tracking

- ORCID is useful because:
  - it’s a **unique**, permanent ID for researchers
  - it has a record closely approximating an “academic CV”
Why ORCID is suitable for outcomes tracking

- ORCID is useful because:
  - it’s a unique, permanent ID for researchers
  - it has a record closely approximating an “academic CV”
  - users can update and *verify* their own information
Why ORCID is suitable for outcomes tracking

● ORCID is useful because:
  ○ it’s a **unique**, permanent ID for researchers
  ○ it has a record closely approximating an “academic CV”
  ○ users can update and *verify* their own information
  ○ it is increasing in uptake/mandates are coming down the pipeline.
Why ORCID is suitable for outcomes tracking

- ORCID is useful because:
  - it’s a **unique**, permanent ID for researchers
  - it has a record closely approximating an “academic CV”
  - users can update and *verify* their own information
  - it is increasing in uptake/mandates are coming down the pipeline.

- It acts as a database of affiliations, and subsequent achievements, likely of interest to those measuring career/research outcomes.
Why ORCID is suitable for outcomes tracking

- ORCID is useful because:
  - it’s a **unique**, permanent ID for researchers
  - it has a record closely approximating an “academic CV”
  - users can update and *verify* their own information
  - it is increasing in uptake/mandates are coming down the pipeline.

- It acts as a database of affiliations, and subsequent achievements, of interest to those measuring career/research outcomes.

- Member organizations can push data to records of users who authorize them, and also retrieve data *made available to them*.
Workflow to get ORCID data

Anyone can pull data from the public API; ORCID Member organizations have the ability to pull, as well as push, data from/to ORCID using the Member API.
Data potentially available in ORCID

Information recorded:

- **Person**
  - *Names* (except other credit names)
  - *Email*
  - *Bio*
  - *Websites/Social Media*
  - *Country*
  - *Keywords*
  - *External IDs*

- **Note:** Demographics Not Included
Data potentially available in ORCID

Information recorded:

- Activities
  - Education
  - Employment
  - Funding
  - Works
  - Distinctions, Invited Positions, Qualifications, Service, Membership
  - *(Journal Article) Peer Review; Research Resources*
Using ORCID to look at a small dataset

- Working with Dr. Alberto Roca, looking at career outcomes for people in the DiverseScholar Doctoral Directory
Using ORCID to look at a small dataset

- Working with Dr. Alberto Roca, looking at career outcomes for people in the DiverseScholar Doctoral Directory

- Database of grads and postdocs from underrepresented groups that includes demographic information and articulation of career intention
Using ORCID to look at a small dataset

- Working with Dr. Alberto Roca, looking at career outcomes for people in the DiverseScholar Doctoral Directory.

- Database of grads and postdocs from underrepresented groups that includes demographic information and articulation of career intention.

- I searched for ORCID for everyone in database; also LinkedIn search; where possible also bringing in “citation” e.g. lab website.
Using ORCID to look at a small dataset

% ORCID\textup{s} vs. Year

\begin{tikzpicture}
    \begin{axis}[
        title={
            \textbf{\% ORCID\textup{s} vs. Year}
        },
        xlabel={Year},
        ylabel={\% ORCID\textup{s}},
        xmin=2015, xmax=2019,
        ymin=0, ymax=100,
        ytick={0, 25, 50, 75, 100},
    ]
    \addplot [color=blue, mark=x] coordinates {
        (2015, 50)
        (2016, 60)
        (2017, 70)
        (2018, 90)
        (2019, 70)
    };
\end{axis}
\end{tikzpicture}
Using ORCID to look at a small dataset

% ORCID vs. Year

% ORCID

2015 2016 2017 2018 2019

Year

All faculty
Using ORCID to look at a small dataset

Overall, 47% of 2015-2019 database members are currently in faculty positions

56% of those with ORCIDs are in faculty positions

(i.e. ORCID is more useful for looking at faculty)
Initial findings working with small public ORCID datasets

Kinds of Information most generally *available* and somewhat *standardized*, for tracking interest:

- Current location and job title
  - Dates
Initial findings working with small public ORCID datasets

Kinds of Information most generally *available* and somewhat *standardized*, for tracking interest:

- Current location and job title
  - Dates
- “Works” e.g. publications (after job info, number of single publication entries (mandate?))
  - Dates
  - Venue
  - Title
Initial findings working with small public ORCID datasets

Kinds of Information most generally *available* and somewhat *standardized*, for tracking interest:

- Current location and job title
  - Dates
- “Works” e.g. publications (after job info, number of single publication entries (mandate?))
  - Dates
  - Venue
  - Title
- Funding
  - Source
  - Amount
  - Title
  - Description
Initial findings working with small public ORCID datasets

ORCID is particularly useful for disambiguation

Currently a reasonable amount of data to work with, potentially more in member API
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
- Variations on data availability, based on what person set as private/public
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
- Variations on data availability, based on what person set as private/public
- Variations on current location e.g. move to LinkedIn
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
- Variations on data availability, based on what person set as private/public
- Variations on current location e.g. move to LinkedIn
- While data is more likely to be verified by author and *correctly attributed*, some may be missing purely by *omission*, or have inaccuracies
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
- Variations on data availability, based on what person set as private/public
- Variations on current location e.g. move to LinkedIn
- While data is more likely to be verified by author and *correctly attributed*, some may be missing purely by *omission*, or have inaccuracies
  - Dependent on how often author maintains account
Initial findings working with small public ORCID datasets

Caveats

- No emails (publicly)
- Variations on data availability, based on what person set as private/public
- Variations on current location e.g. move to LinkedIn
- While data is more likely to be verified by author and *correctly attributed*, some may be missing purely by *omission*, or have inaccuracies
  - Dependent on how often author maintains account
  - However, may be likely to change with increasing focus on externally verified info
Initial findings working with small public ORCID datasets

Opportunities

- Member orgs have access to potentially more data through member API
  - Last update date very recent i.e. same day (likely through API connections) - which is public, even if no other information is, suggesting more data likely to be available from member API.
Initial findings working with small public ORCID datasets

Opportunities

- Member orgs have access to potentially more data through member API
  - Last update date very recent i.e. same day (likely through API connections) - which **is public**, even if no other information is, suggesting more data likely to be available from member API.

- Communication efforts could drastically increase the information available to member organizations
Initial findings working with small public ORCID datasets

Opportunities

- Member orgs have access to potentially more data through member API
  - Last update date very recent i.e. same day (likely through API connections) - which is public, even if no other information is, suggesting more data likely to be available from member API.
- Communication efforts could drastically increase the information available to member organizations
- For various reasons, ORCID data is only likely to continue to get *better*
Initial findings working with small public ORCID datasets

Opportunities

● Member orgs have access to potentially more data through member API
  ○ Last update date very recent i.e. same day (likely through API connections) - which *is public*, even if no other information is, suggesting more data likely to be available from member API.

● Communication efforts could drastically increase the information available to member organizations

● For various reasons, ORCID data is only likely to continue to get *better*

● Opportunities to push “service” etc into unused categories (perhaps standardize?)
Future Work

Suggested future directions for the HRA community

- Move towards long-term reliance on ORCID data
Future Work

Suggested future directions for the HRA community

- Move towards long-term reliance on ORCID data
- Work with developer to set up what your organizations need
Future Work

Suggested future directions for the HRA community

- Move towards long-term reliance on ORCID data
- Work with developer to set up what your organizations need
- Work with ORCID on specific outcomes tracking needs
Future Work

Suggested future directions for the HRA community

- Move towards long-term reliance on ORCID data
- Work with developer to set up what your organizations need
- Work with ORCID on specific outcomes tracking needs
  - what is your ideal dataset?
Future Work

Suggested future directions for the HRA community

- Move towards long-term reliance on ORCID data
- Work with developer to set up what your organizations need
- Work with ORCID on specific outcomes tracking needs
  - what is your ideal dataset?
  - advocate for what is needed, that they could deliver
General Personal Thoughts on Use of Evaluation Tools
General Personal Thoughts on Use of Evaluation Tools

Tools that allow you to understand where the data is coming from/how it is aggregated in a transparent fashion should be a priority - so that you can be fully aware of the caveats.

Greater focus on data about *people* in concert with e.g. publications data will allow for appreciation of context and (dis)advantages of different groups.
Acknowledgements and Questions