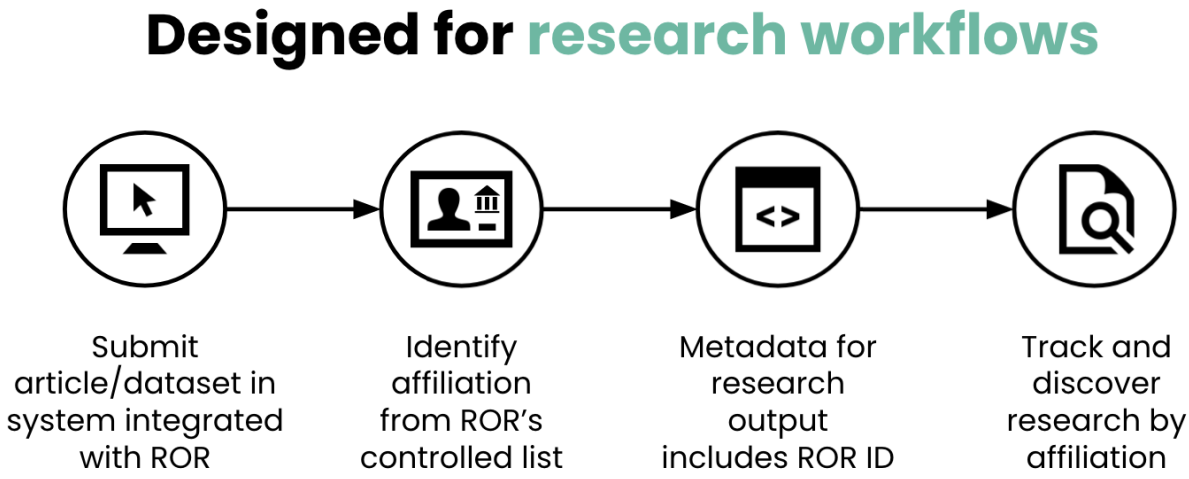
ROR for Repositories

Context

Repositories can use the Research Organization Registry (ROR) to help ensure that organization information is clean and consistent within and between information systems. ROR IDs are particularly useful for identifying 1) which organization(s) an author/creator of a research output is affiliated with, and 2) which organization(s) have funded a piece of research.

Using ROR IDs for organizations not only helps to streamline internal workflows and automate internal reporting but also helps the wider community gain insight into institutional associations with research. As a free and open identifier for research organizations with CC0 metadata, ROR is ideal for inclusion in DOI metadata submitted to DataCite and Crossref.

This document is meant to be a guide written by and for repository managers, developers, and metadata specialists to help them make decisions about how to implement ROR.



Integrating ROR into submission forms

General advice

- Consult https://ror.readme.io/docs/create-ror-powered-typeaheads for more detailed instructions.
- Where possible, use ROR IDs for author/creator/contributor affiliations, for organizations considered to be creators/authors, AND for funder organizations in your submission forms and workflows.
- Never require the ROR ID. If the field is required, allow the user to enter free text so that they can enter organizations that are not in ROR. Do not prompt the user to request a ROR ID, since such requests can take weeks to process and may be denied.

Submission form Dos and Don'ts

- **Do** allow users to choose organizations by name rather than by ROR ID.
- **Don't** ask users to know, search for, or enter their organization's ROR ID.
- **Do** use a a ROR-powered typeahead to let users select organizations.
- **Do** capture units, subunits, and departments of organizations in separate fields. Best practice is to use a ROR-powered typeahead only for high-level organizations and to provide separate free-text fields for organizational units such as university departments that do not have a ROR ID.

- **Do** allow users to choose multiple organizations, since many people are affiliated with multiple organizations and many research outputs are supported by multiple funders.
- **Do** display supplemental information to the user to enable them to choose the correct organization. In addition to the name of the organization, consider showing such information as aliases, acronyms, labels, location information, and URL.
- **Do** allow users to enter organization names in a case-insensitive manner (e.g., "NOA" and "Noa" should produce the same results), but consider showing results that match entire acronyms before those that match part of a name (e.g., "National Observatory of Athens (NOA)" before "Noah's Ark Children's Hospital").
- **Do** consider whether it is necessary or helpful to the user to display the ROR ID. If you do decide to display the ROR ID, please follow the ROR ID Display Guidelines.

Implementation advice

- Rather than making calls to the live ROR API, consider using a local copy of the ROR dataset or running a local instance of the ROR API in Docker.
    - The live ROR API always has the most current data and is generally reliable, but as publicly-available, community-shared open infrastructure traffic (and therefore response time) can vary. The ROR API's rate limit is a maximum of 2000 requests in a 5-minute period, and API traffic can be quite heavy at popular times, like midnight UTC.
    - Note that it is not possible to retrieve the entire ROR dataset from the API as Elasticsearch has an inherent upper limit of 10,000 results. If you need to obtain the entire ROR dataset, use the data dump.
    - Storing ROR data locally for your system to query is inexpensive and stable, but since the ROR dataset is updated at least monthly (more often twice a month), local ROR data may get out of date quickly. Data dumps can be retrieved programmatically using the Zenodo API.
    - Installing the ROR API locally in Docker is inexpensive and stable, but it relies on multiple dependencies such that hardware and software versioning may become an issue.
- If you use the ROR API, check its heartbeat at https://api.ror.org/heartbeat.
- Retrieve only records whose status is "active".
- Recommended practice is to capture and store the ROR ID as a complete URL in the format `https://ror.org/02mhbdp94`. However, the ROR API will also recognize forms such as `ror.org/02mhbdp94` or `02mhbdp94`, so these forms of the ROR ID may also be used.
- Do not require users to enter an organization with a ROR ID even if the organization field is required; allow the user to enter free text that overwrites the ROR suggestion in form fields.
- After a user chooses an organization from the ROR-powered list, don't allow them to edit the name. Instead, prompt them to choose a new organization from the ROR list *or* enter their own.

- Make organization fields repeatable, since many people are affiliated with multiple organizations.
- Match user-entered data against all name-related fields in ROR metadata: 'name', 'aliases', 'acronyms', and 'labels'.
- Consider filtering the list of organizations displayed to the user based on context, such as the user's location (browser geolocation, location info entered in other fields on the same form), email domain, and/or organization type.
- Consider populating other fields in your form automatically, using data from the ROR record of the organization selected by the user.
- If you retrieve location metadata from ROR, be aware of differences between fields.
    - In the current ROR schema (version 1) there are multiple fields for country names: the `country` field and the `country.country_name` field. Values in both should in general be the same, but the `country.country_name` data comes from the external Geonames API.
    - For cities, there is only one field to pull from: addresses.geonames_city.city, which also comes from Geonames.
    - In the U.S., developers might also wish to show the U.S. state, which can be drawn from the `addresses.geonames_city.geonames_admin1.name` field.

User journey

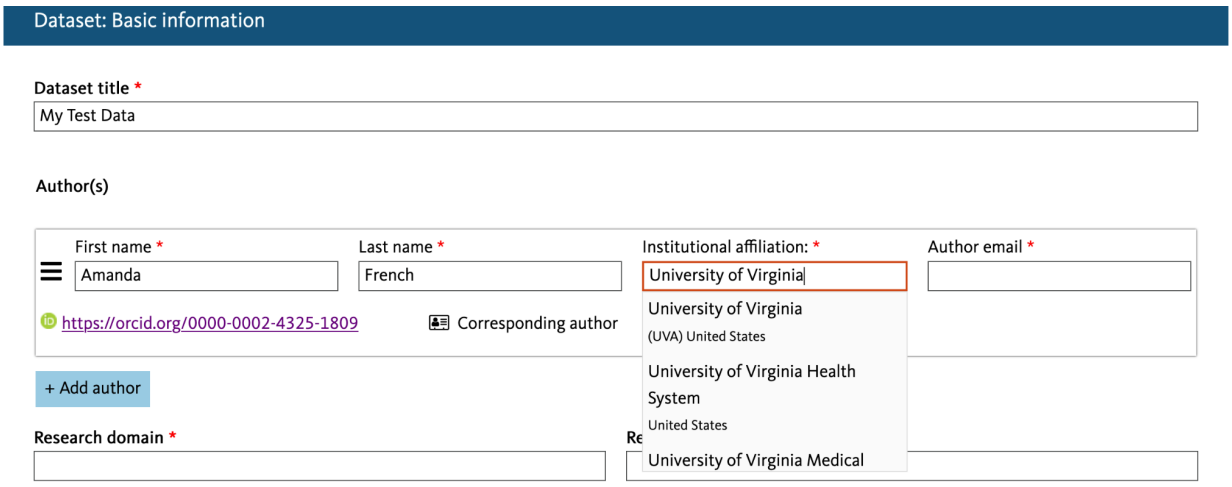
- User sees an Institution Name text entry field in the form.
- User begins to type an institution name in the Institution Name field and a dropdown of suggestions appears.
    - The first line of each suggestion includes the institution's name.
    - Additional information displayed in the suggestion includes the institution's country, and, optionally, its city. U.S. implementations may also wish to include the institution's state.
    - Further information displayed in the suggestion includes acronyms and alternative names for the institution, the institution's name in other languages, and, optionally, the institution's website.
- User can select an institution from the dropdown list, which autopopulates the Institution Name field with ‘name’ from ROR metadata.
    - If a user selects an institution from the dropdown list, the user cannot edit the Institution Name field.
    - The user can delete the institution name selected from the dropdown. This action also deletes the ROR ID.
- User can type an institution name into the text field rather than choosing one from the dropdown list. This results in text-only institutional name data unmatched to a ROR ID.

Examples and screenshots

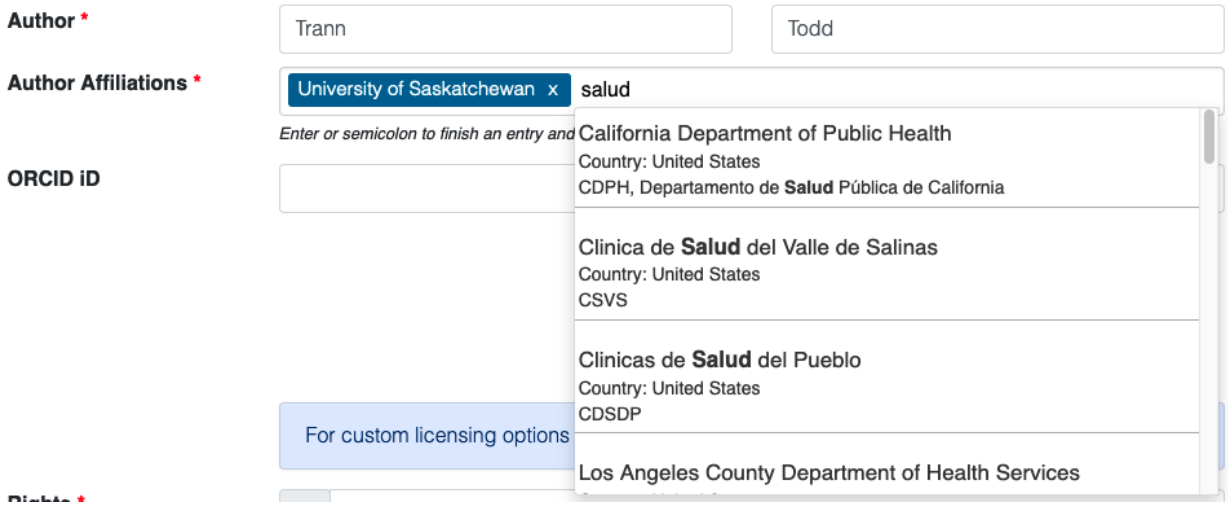
- "ROR-ing Together with Dryad" - https://ror.org/blog/2019-07-10-ror-ing-together-with-dryad/

- Federated Research Data Repository (FRDR) - https://demo.frdr-dfdr.ca
- "Add Research Institution Identifiers with ROR" - https://info.orcid.org/add-research-institution-identifiers-with-ror/
- ROR Typeahead Demos - https://ror-community.github.io/ror-typeahead-demos/
- ROR Typeahead Demos codebase - https://github.com/ror-community/ror-typeahead-demos

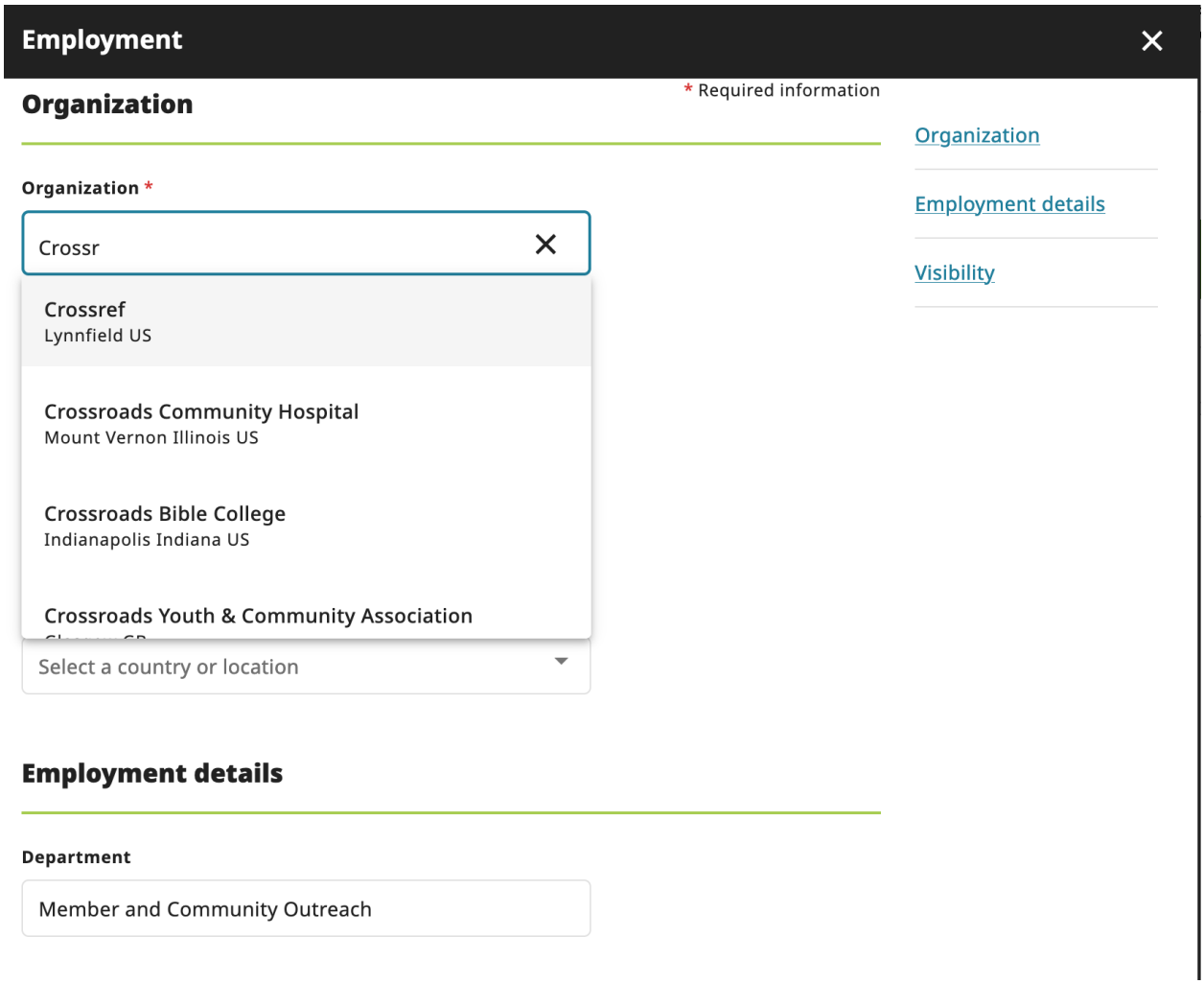
Dryad Screenshot



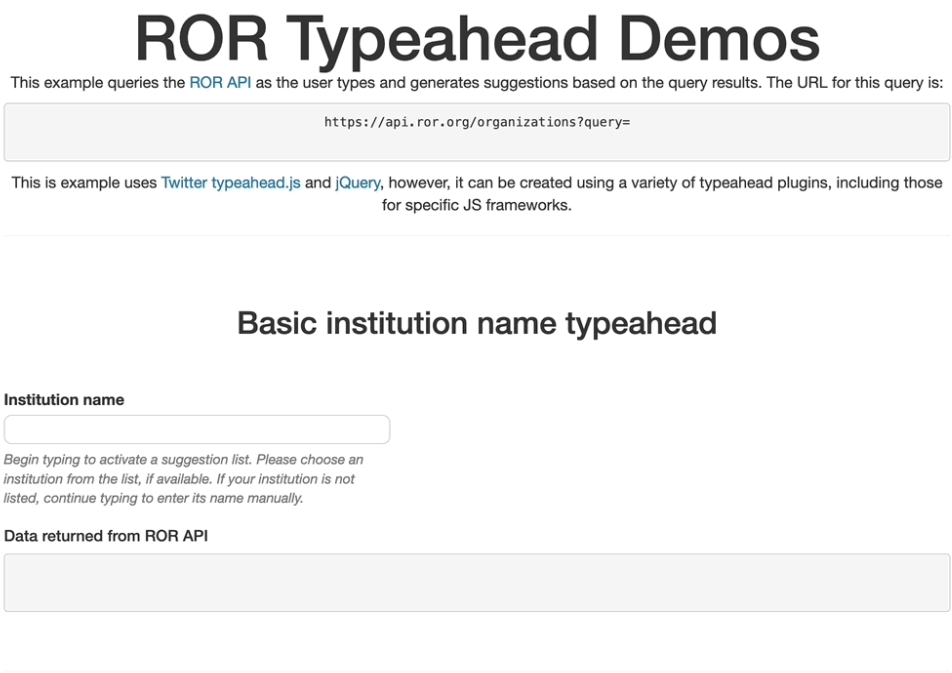
FRDR Screenshot



ORCID Screenshot



ROR Typehead Demo Screenshot



Matching existing metadata to ROR IDs

General advice

- Consult https://ror.readme.io/docs/match-organization-names-to-ror-ids for detailed instructions on matching text strings to ROR IDs.
- Consult https://ror.readme.io/docs/map-other-organization-id-types-to-ror for detailed instructions on matching other organizational identifiers to ROR IDs.
- If you have both organizational identifiers and organization names in your data, you will get quicker, more accurate results by matching IDs than by matching names.
- Remember that no automatic tool will ever be able to match text strings (organization names) to ROR IDs with 100% accuracy, although high rates (80%-90%) of correct matches are possible with automated tools. Best practice is to add a layer of human review and quality assurance after automatic matching.
- Consider whether to use an existing script or tool that is designed for someone else's data (such as those listed in the ROR documentation) or a script or tool you write yourself that takes features of your data into consideration.
- Make a plan for how your metadata will or will not change when organizations merge, shut down, or change names. If you change your metadata internally, best practice is to update DOI metadata as well.

Requesting changes and additions to ROR records

- New ROR records or updates to individual ROR records can be requested via a form https://curation-request.ror.org.
- For bulk requests to correct metadata or add records for multiple organizations, please email registry@ror.org.
- New versions of the ROR registry are currently released every 2-4 weeks. Consider how often you need to refresh your ROR data if it is stored locally.

Adding ROR IDs to DOI metadata

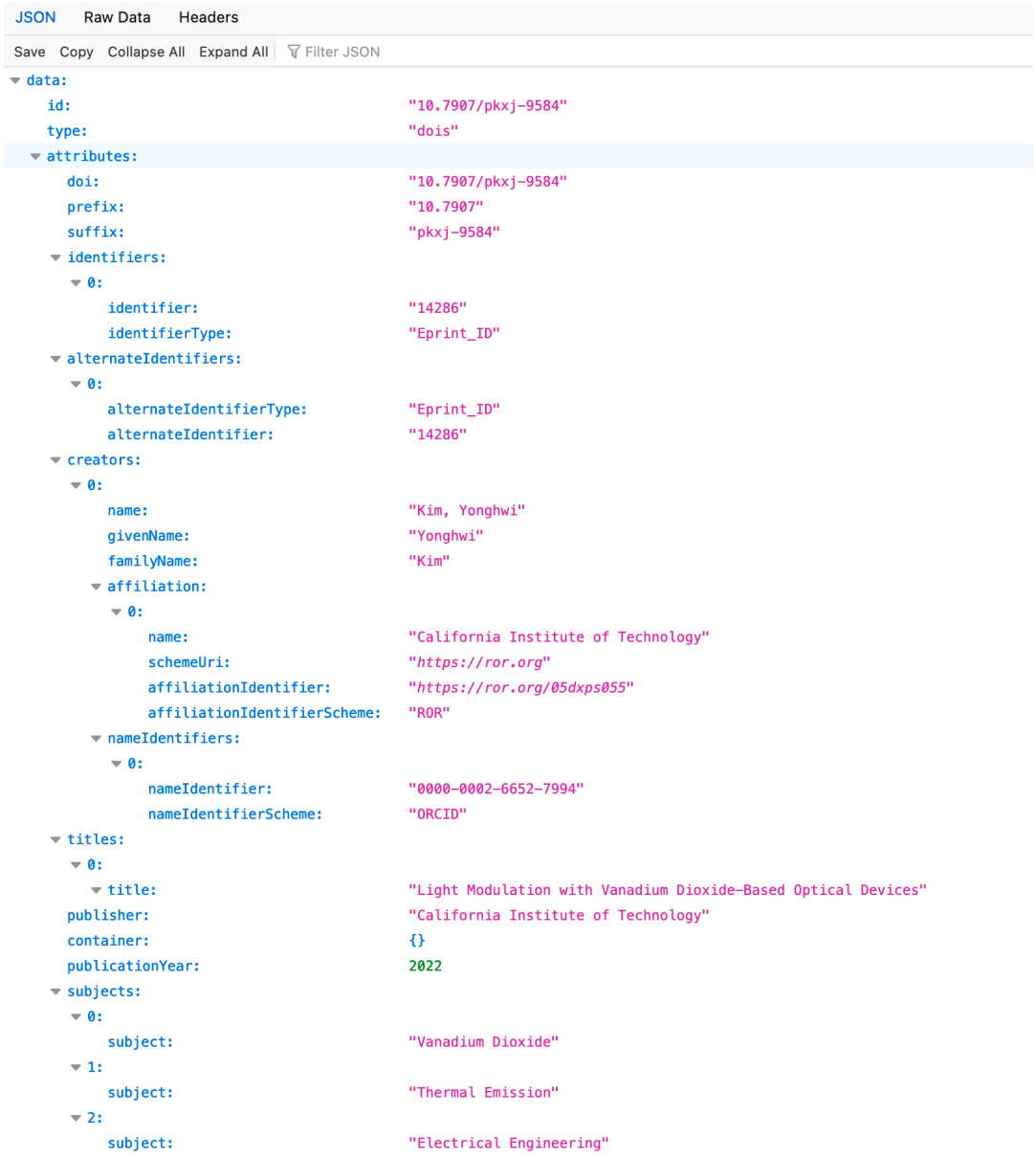
General advice

- When you register DOIs with DataCite, Crossref, or any other DOI registrar, include ROR IDs in the metadata.
- If possible, include ROR IDs for author/creator affiliations, organizations as authors/creators, and funding organizations.
- The forthcoming DataCite metadata schema, version 4.5, will also support ROR IDs for publishers. Consider using ROR IDs for publishers in DOI metadata.
- If you update existing organization information in your system to include ROR IDs, be sure to update your existing DOI metadata with your registrar as well.
- Make sure that the institution name is actually there when you send ROR IDs for subunits to DataCite. You can repeat the affiliation field for subunits.

Sending ROR IDs to DataCite

- ROR documentation: Include ROR IDs in DataCite DOIs
- DataCite documentation: Connecting to Organizations
- DataCite documentation: What is the recommended format for including nameidentifiers in DataCite metadata?
- DataCite documentation: Can I see more detailed information in the REST API?

Sample DataCite record with ROR ID



Sample DataCite API and Commons queries

- DataCite Commons: California Institute of Technology: https://commons.datacite.org/ror.org/05dxps055
- ROR IDs in creator/contributor affiliations in DataCite: https://api.datacite.org/dois?query=(creators.affiliation.affiliationIdentifierScheme:ROR+OR+contributors.affiliation.affiliationIdentifierScheme:ROR)&affiliation=true
- ROR IDs in funding references in DataCite: https://api.datacite.org/dois?query=fundingReferences.funderIdentifierType:ROR

Sending ROR IDs to Crossref

- ROR documentation: Include ROR IDs in Crossref DOIs
- Crossref documentation: Affiliations and ROR

Sample Crossref record with ROR ID


```
message:
  institution:
    0:
      id:
        0:
          id:                 "https://ror.org/02e3zdp86"
          id-type:            "ROR"
          asserted-by:        "publisher"
        name:                 "Boise State University"
      acronym:
        0:                    "BSU"
      place:
        0:                    "Boise, Idaho, USA"
  indexed:                    {…}
  posted:
    date-parts:
      0:
        0:                    2018
        1:                    4
  group-title:                "Andrus Center for Public Policy Publications and Presentations"
  reference-count:            0
  publisher:                  "Boise State University"
  content-domain:
    domain:                   []
    crossmark-restriction:    false
  short-container-title:      []
  abstract:                   "<p>. While the Treasure …eholder engagement.</p>"
  DOI:                        "10.18122/andrus_center_pubs/1/boisestate"
  type:                       "posted-content"
  created:
    date-parts:
      0:
        0:                    2019
        1:                    4
        2:                    18
    date-time:                "2019-04-18T15:02:53Z"
    timestamp:                1555599773000
  source:                     "Crossref"
  is-referenced-by-count:     0
  title:
    0:                        "Idaho's Water"
  prefix:                     "10.18122"
  author:
    0:
      ORCID:                  "http://orcid.org/0000-0003-1773-6236"
      authenticated-orcid:    true
      given:                  "Kendra"
      family:                 "Kaiser"
      sequence:               "first"
      affiliation:
        0:
          id:
            0:
              id:             "https://ror.org/02e3zdp86"
              id-type:        "ROR"
              asserted-by:    "publisher"
            name:             "Boise State University"
```

Sample Crossref API queries

- ROR IDs in the `institution` element for dissertations in Crossref: https://api.crossref.org/works?filter=has-ror-id:t,type-name:Dissertation
- DataCite API query by ROR ID: https://api.datacite.org/dois/10.7907/pkxj-9584?affiliation=true

Resources

- ROR Documentation - https://ror.readme.com
- ROR Demos, Tutorials, and Presentations - https://www.youtube.com/@researchorgs/playlists
- ror-utilities repository for third-party ROR scripts and code - https://github.com/ror-community/ror-utilities
- Request a ROR feature or report a ROR bug on the ROR roadmap - https://github.com/ror-community/ror-roadmap/issues
- Request a new ROR ID or a change to a ROR record - https://curation-request.ror.org