Datenkuratierung, Entwicklung eines Metadatenschemas & Abfragemöglichkeiten für ROR-IDs

KIM Workshop 2021

Paul Vierkant, DataCite, https://orcid.org/0000-0003-4448-3844

Meeting agenda

- What is (in) ROR?
- 2020 highlights/2021 preview
- Spotlight on: adoption & integration
- Spotlight on: curation

Kennen Sie ROR?

A/ Noch nie gehört
B/ Grundsätzlich ja
C/ Längst implementiert

ROR

What is ROR and who is the ROR Community?

ROR

What is ROR and why does it exist?

"A community-led registry of open, sustainable, usable, and unique identifiers for every research organization in the world"

- Focused specifically on affiliations in scholarly research
- Developed with, by, and for the community
- Open and noncommercial
- Infrastructure designed to connect organizations to researchers and research outputs

Current service model

- Registry of IDs and metadata
- Open tools for interacting with the registry
  - API (search, retrieve, affiliation match)
  - Search page
  - Data dumps (approximately quarterly)
  - Reconciler

Current structure

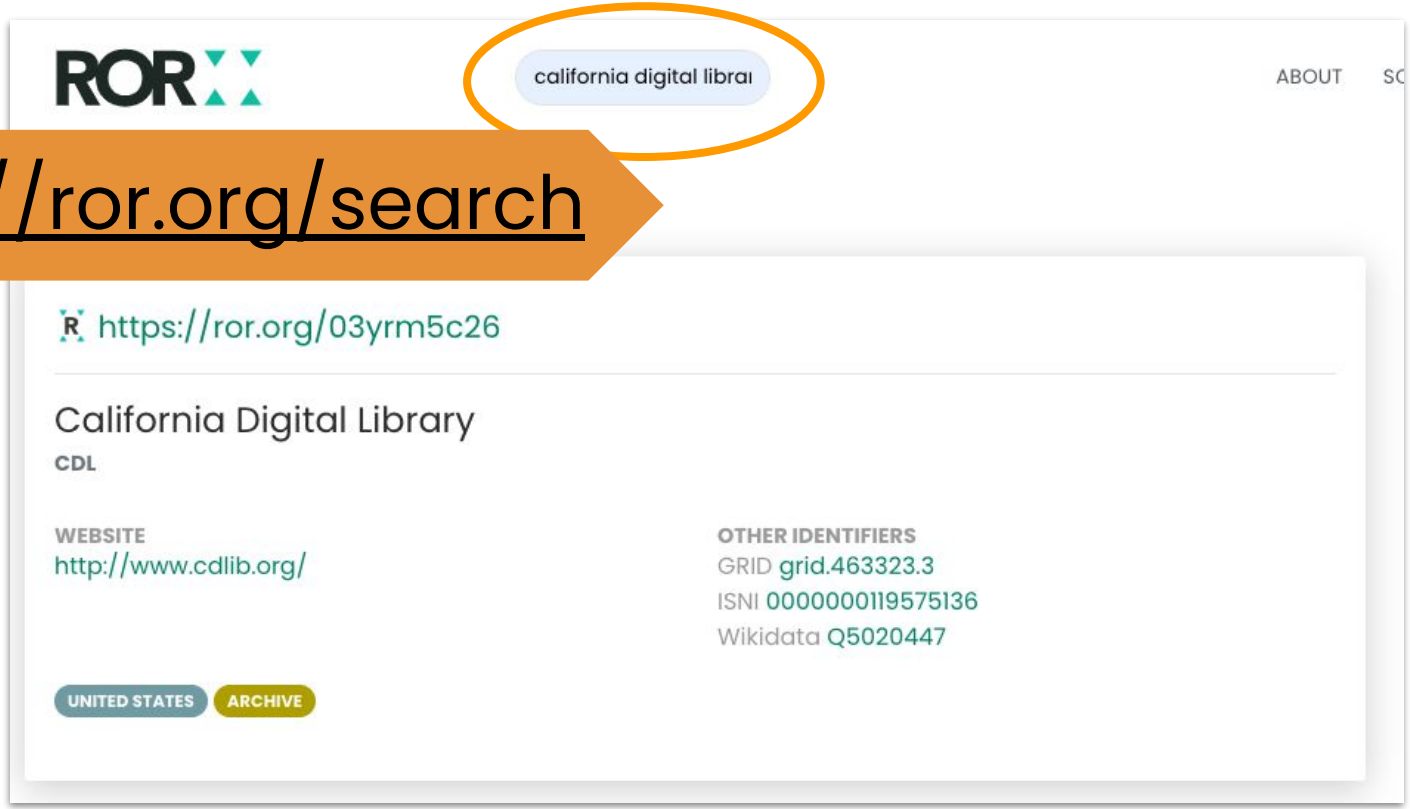
- ROR is **not an organization** - it is a **multi-organization collaboration**
  - Operating orgs + steering group
  - Dedicated and ad hoc staff from operating orgs
- **Community engagement** through advisory groups
  - General community group - open to all
  - Adoption group - those working on integrations
  - Curation group - those advising on curation

ROR sustainability model

- Low overhead, focused scope
- Early and broad community engagement and adoption
- In startup phase (2019-2022), resource ROR with a mixed funding model (in-kind support from ROR operating orgs, community donations, grants)
- Optional paid service tier to launch in 2022 (registry data will always be open and free)

(How do I see) what's in ROR?

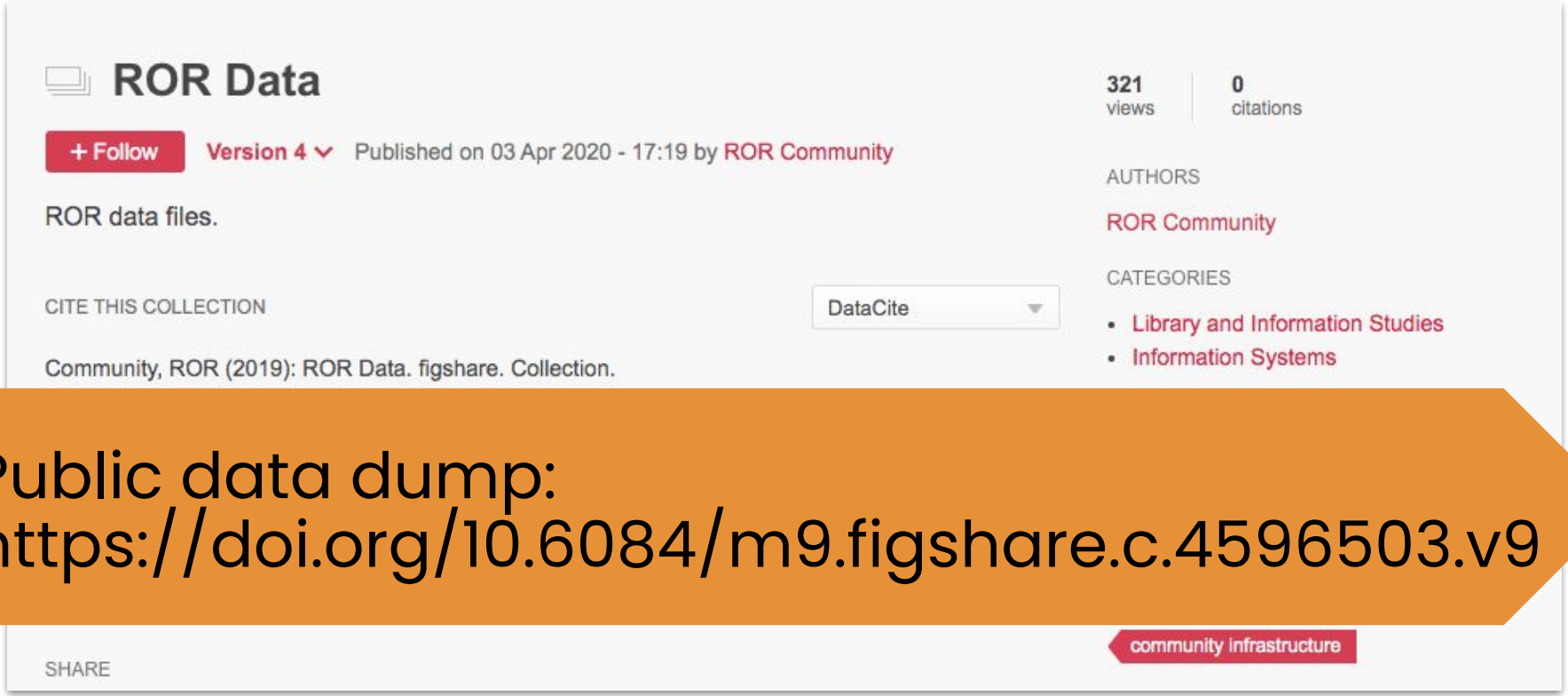
https://ror.org/search



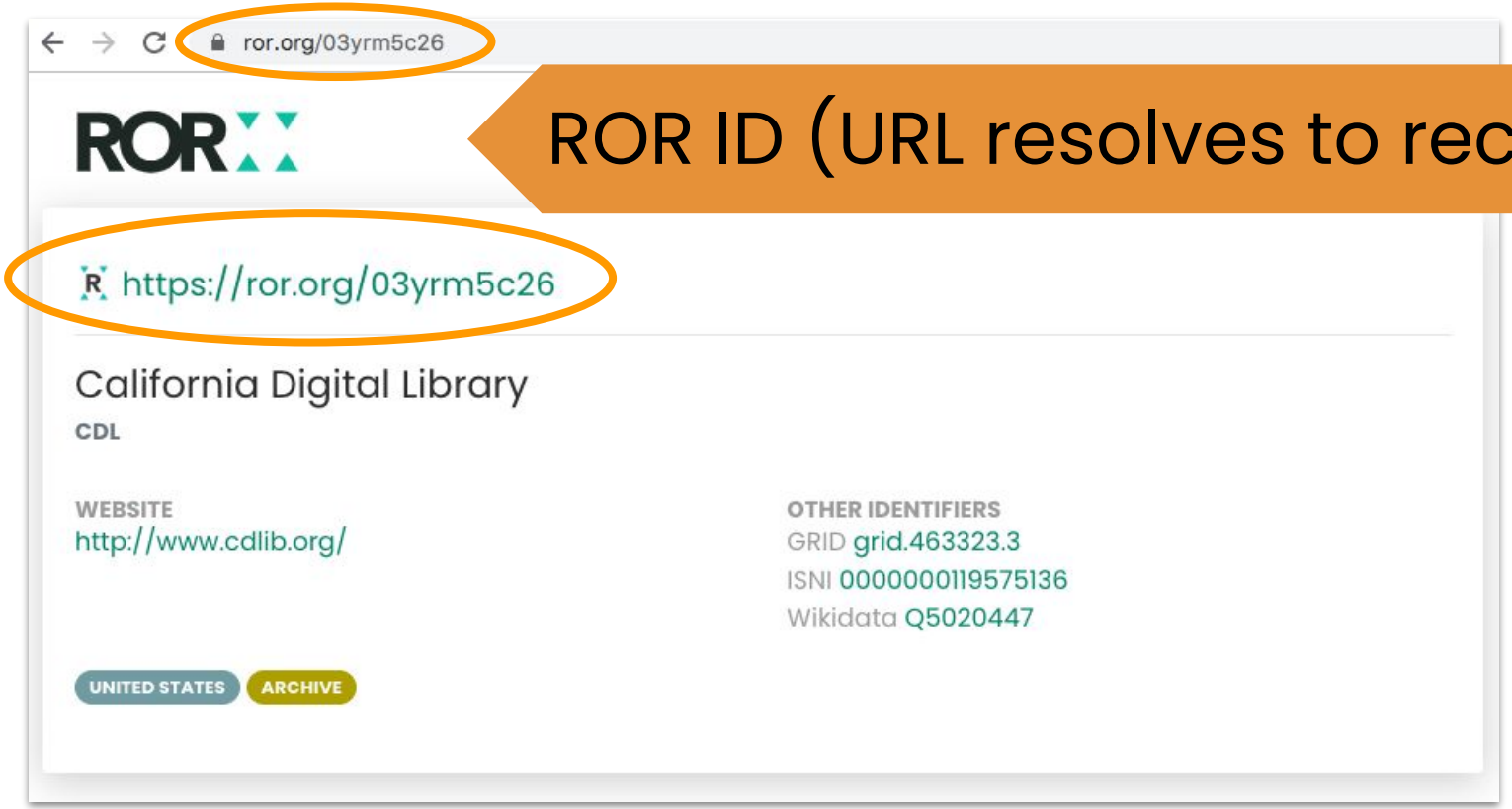
ROR API:
https://api.ror.org/organizations

ROR

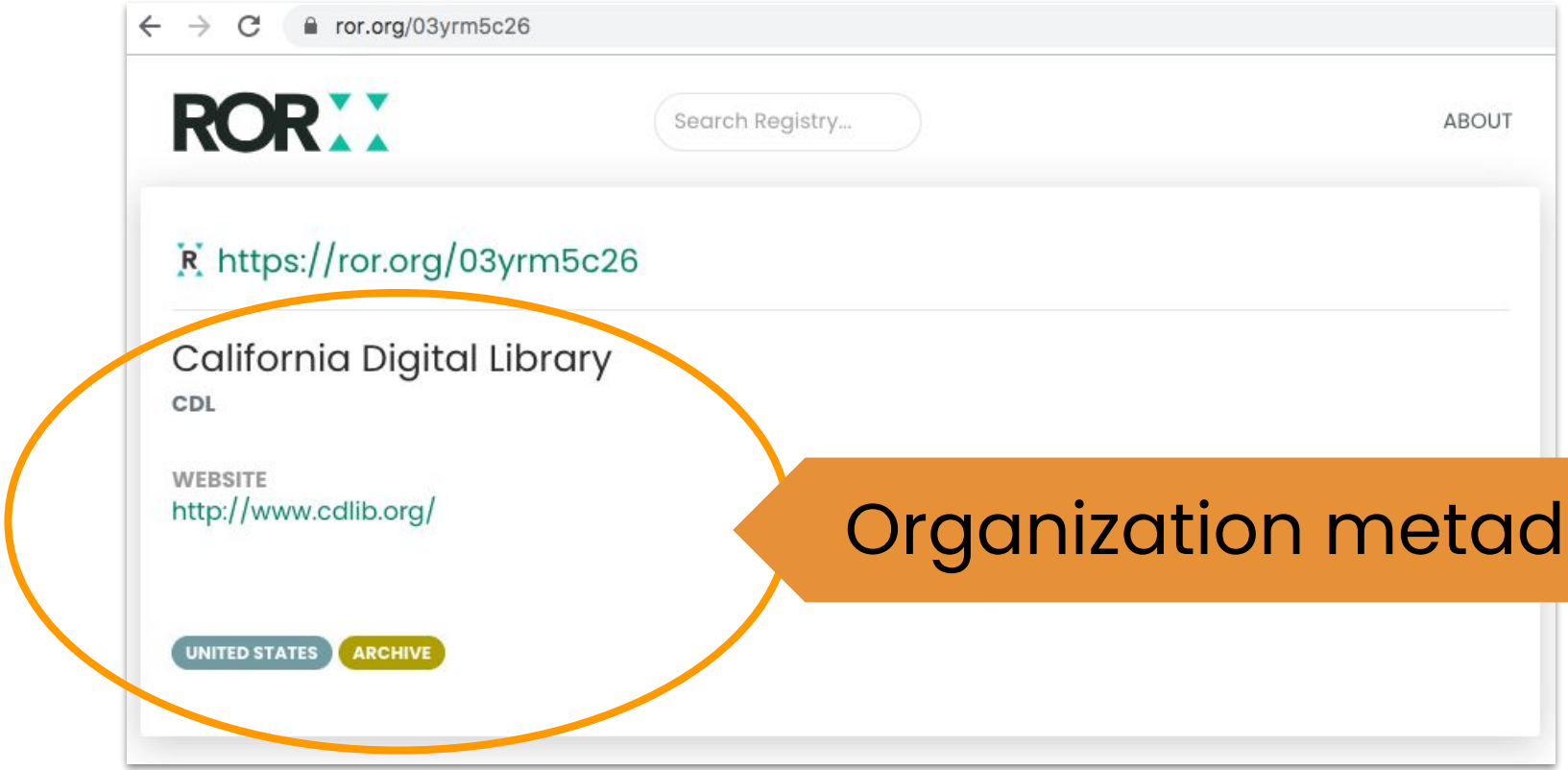
@ResearchOrgs



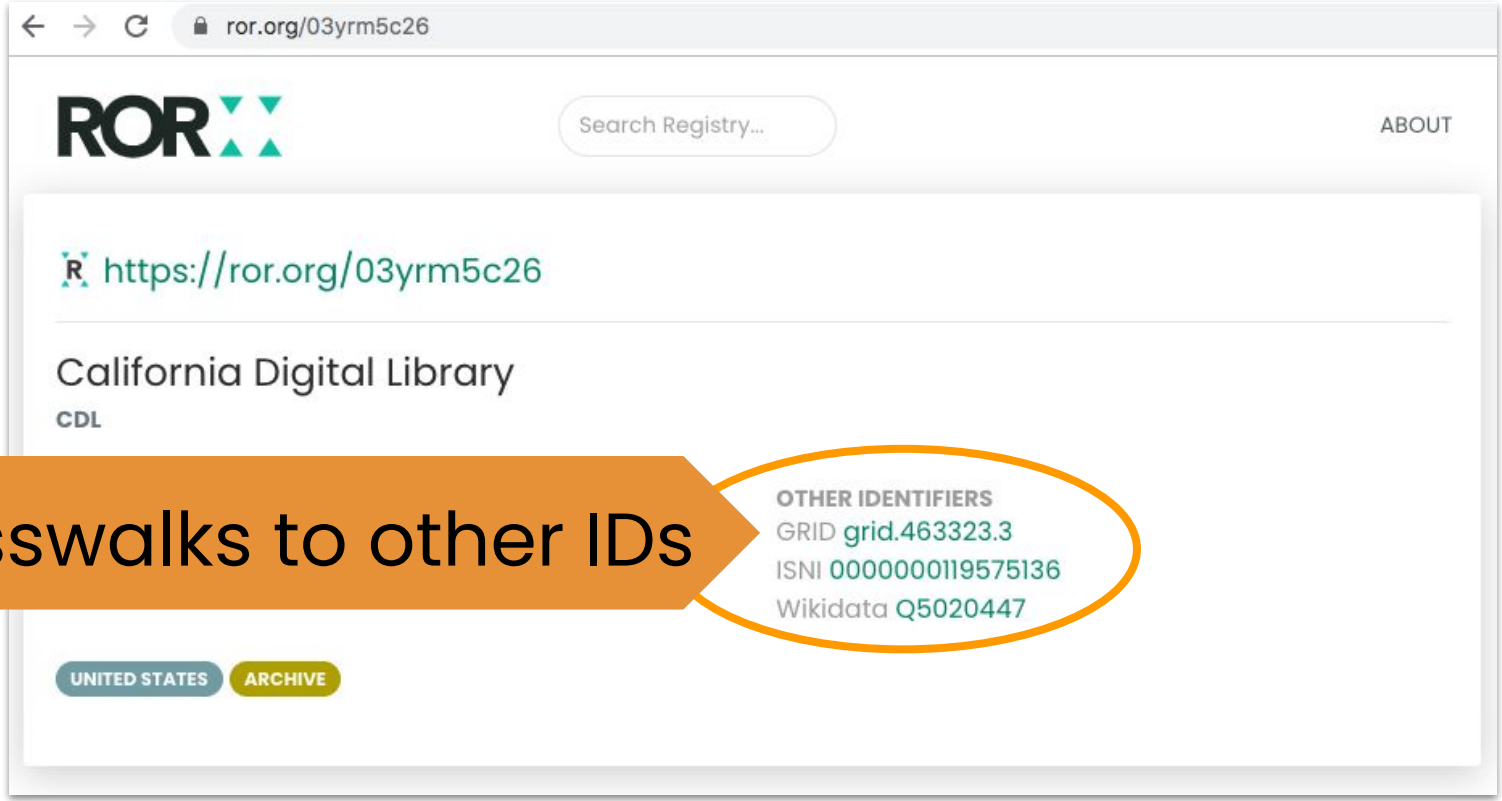
Public data dump:
https://doi.org/10.6084/m9.figshare.c.4596503.v9



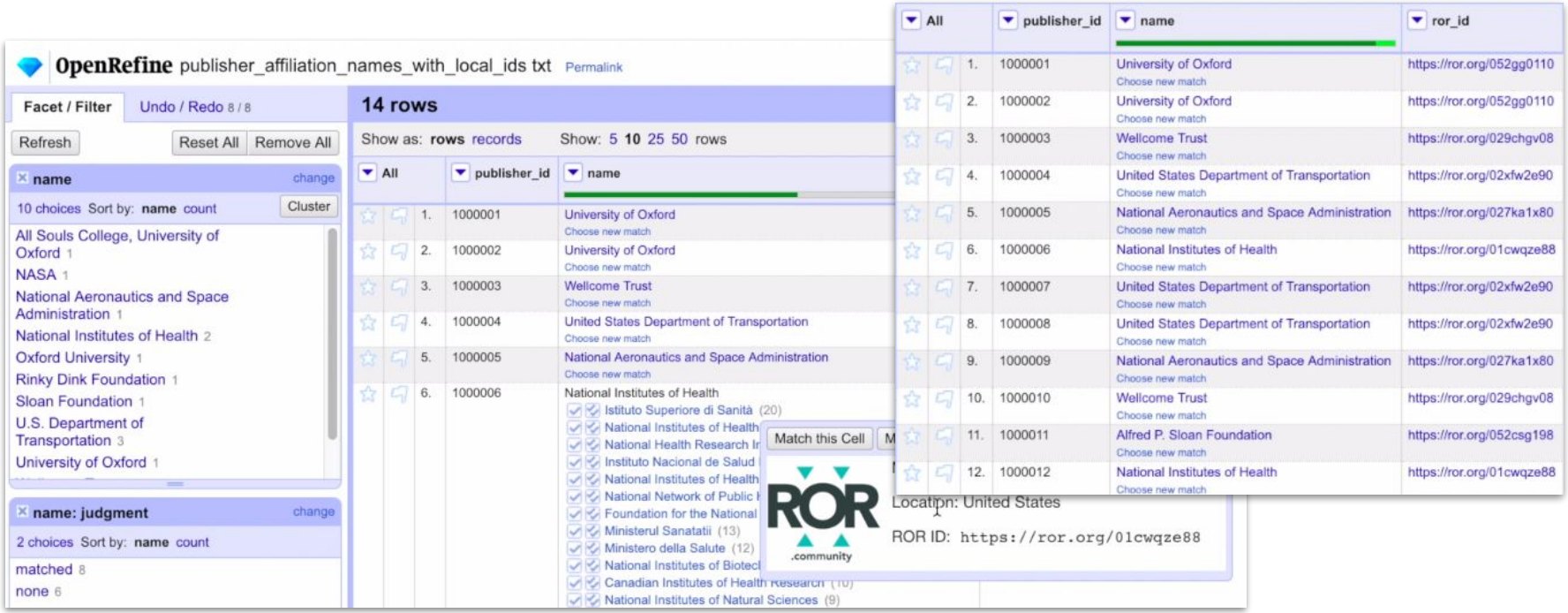
ROR ID (URL resolves to record)


ror.org/03yrm5c26
ROR
Search Registry...
ABOUT
https://ror.org/03yrm5c26
California Digital Library
CDL
WEBSITE
http://www.cdlib.org/
UNITED STATES
ARCHIVE
Organization metadata

Crosswalks to other IDs



ROR reconciler



ROR

@ResearchOrgs

Affiliation matching

ROR

@ResearchOrgs

Metadata schema

Main format

Key	Type	Required	Description
id	String	True	A unique ID for the institute
name	String	True	The name typically used to refer to the institute.
status	String	True	For an active institute, this is always set to "active"
aliases	[String]	False	A list of other names the institute is known as
acronyms	[String]	False	A list of short acronyms the institute is known as (e.g. MRC for the Medical Research Council)
types	[String]	False	A list of types describing the institute. For an explanation of valid values see our policies
wikipedia_url	String	False	URL of the wikipedia page for the institute
email_address	String	False	A contact email address for the institute
ip_addresses	[String]	False	IP addresses known to belong to the institute
established	Integer	False	The year the institute opened, CE
close	Integer	False	The year the institute closed, CE
links	[String]	False	An array of URLs linking to things like the homepage for the institute
relationships	[Relationships]	False	Any relationships the institute has to others. See our policies for more details
external_ids	External ID	False	Other IDs known to refer to the institute. See our policies for more details
addresses	[Address]	True	An array of addresses associated with the institute, address format described
labels	[Labels]	True	The name of the institute in different languages

Relationship format

Key	Type	Required	Description
type	String	True	The relationship type, see our for more details on the types and the meaning of these
label	String	True	The name of the related institute
id	String	True	The unique ID of the related institute

External ID format

Key	Type	Required	Description
External ID Type	External ID Type	True	An object mapping the external ID name to a list of all IDs and a preferred ID if one exists.

External ID Type

Key	Type	Required	Description
preferred	String	True	The preferred ID for this type, or null if none are preferred
all	[String]	True	A list of all IDs in the "External ID Type" format

Address format

Key	Type	Required	Description
city	String	True	The name of the city
country	String	True	The name of the country
country_code	String	True	The ISO 3166-1 alpha-2 code of the country
country_geonames_id	String	True	The GeoNames ID of the country
lat	Float	False	Latitute of the institute
lng	Float	False	Longitude of the institute
line_1	String	False	First line of the address
line_2	String	False	Second line of the address
line_3	String	False	Third line of the address
postcode	String	False	The postcode/zipcode
primary	Boolean	True	If there is more than one address, identifies the main location
state	String	False	The name of the state/region
state_code	String	False	The ISO 3166-2 region code
geonames_city	Geoname	False	The linked GeoNames data

Metadata schema

GeoNames format

Key	Type	Required	Description
id	String	True	The GeoNames ID
city	String	True	The name of the city
geonames_admin1	Region	True	The admin1 region in GeoNames, documented
geonames_admin2	Region	True	The admin2 region in GeoNames, documented
nuts_level1	Region	True	The NUTS level 1 region the GeoNames city is in, documented
nuts_level2	Region	True	The NUTS level 2 region the GeoNames city is in, documented
nuts_level3	Region	True	The NUTS level 3 region the GeoNames city is in, documented

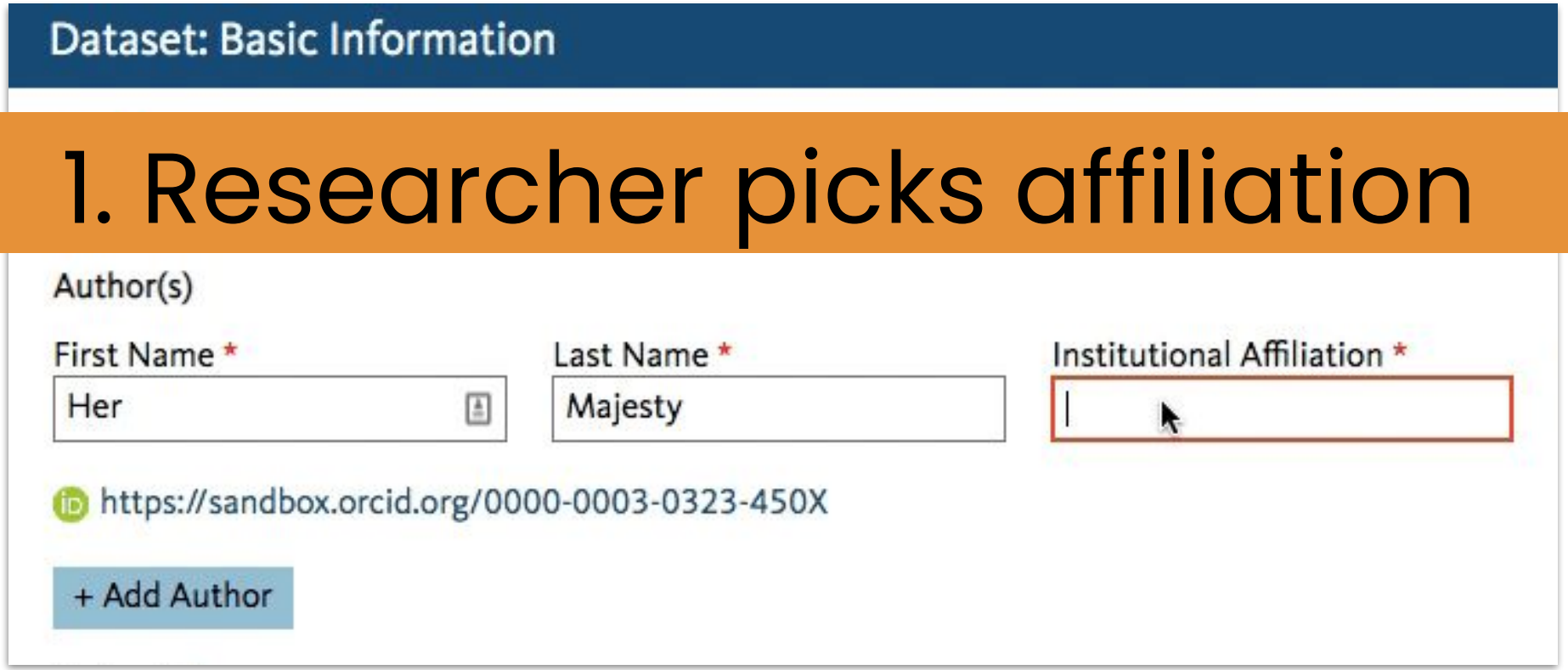
Region format

Key	Type	Required	Description
id	String	True	The ID in the region format
name	String	True	The name of the region
ascii_name	String	False	A preferred ASCII encoded name for the region

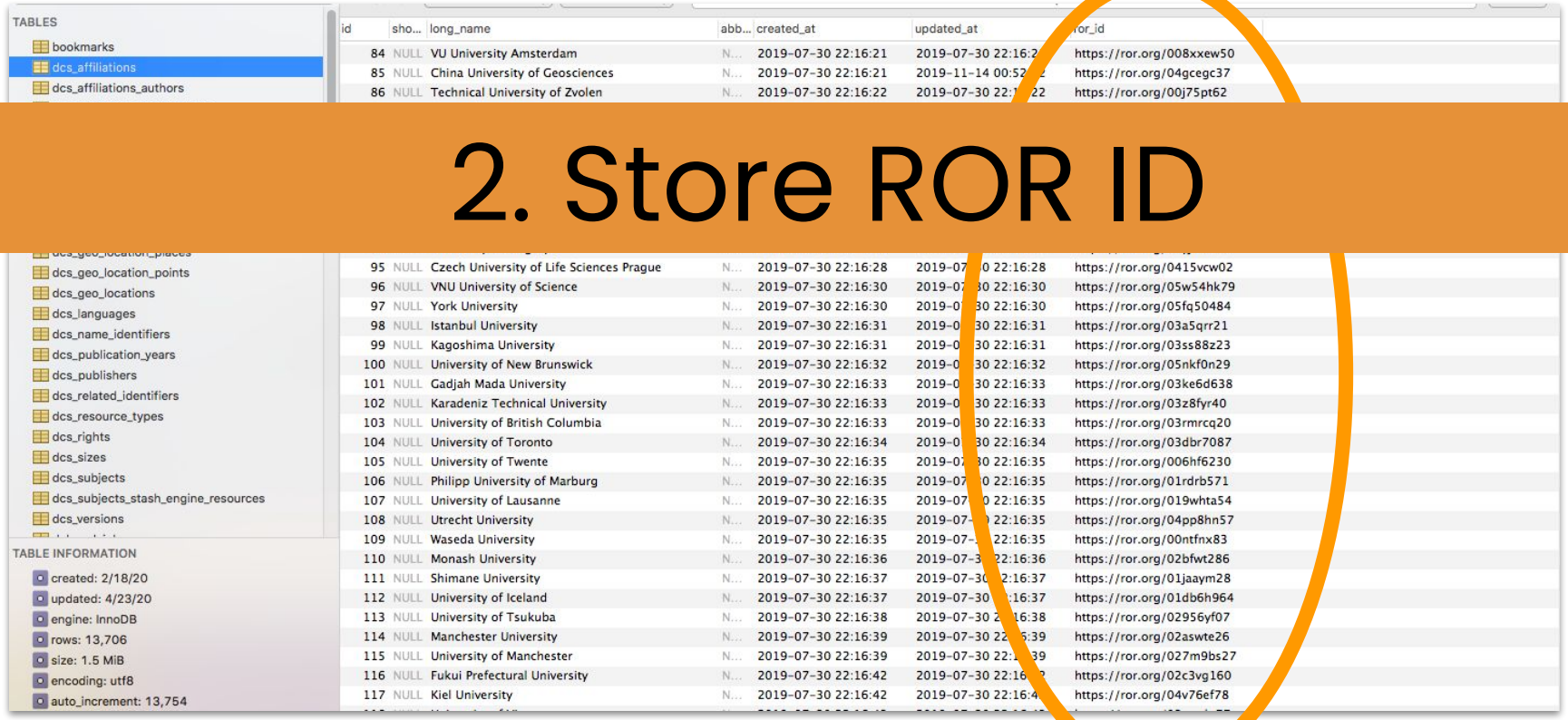
Labels format

Key	Type	Required	Description
label	[String]	True	The institute name in a language variant
iso639	[String]	True	The ISO-639-1 language code

How can ROR work?


Dataset: Basic Information
1. Researcher picks affiliation
Author(s)
First Name *
Her
Last Name *
Majesty
Institutional Affiliation *
https://sandbox.orcid.org/0000-0003-0323-450X
+ Add Author

2. Store ROR ID



ROR

@ResearchOrgs

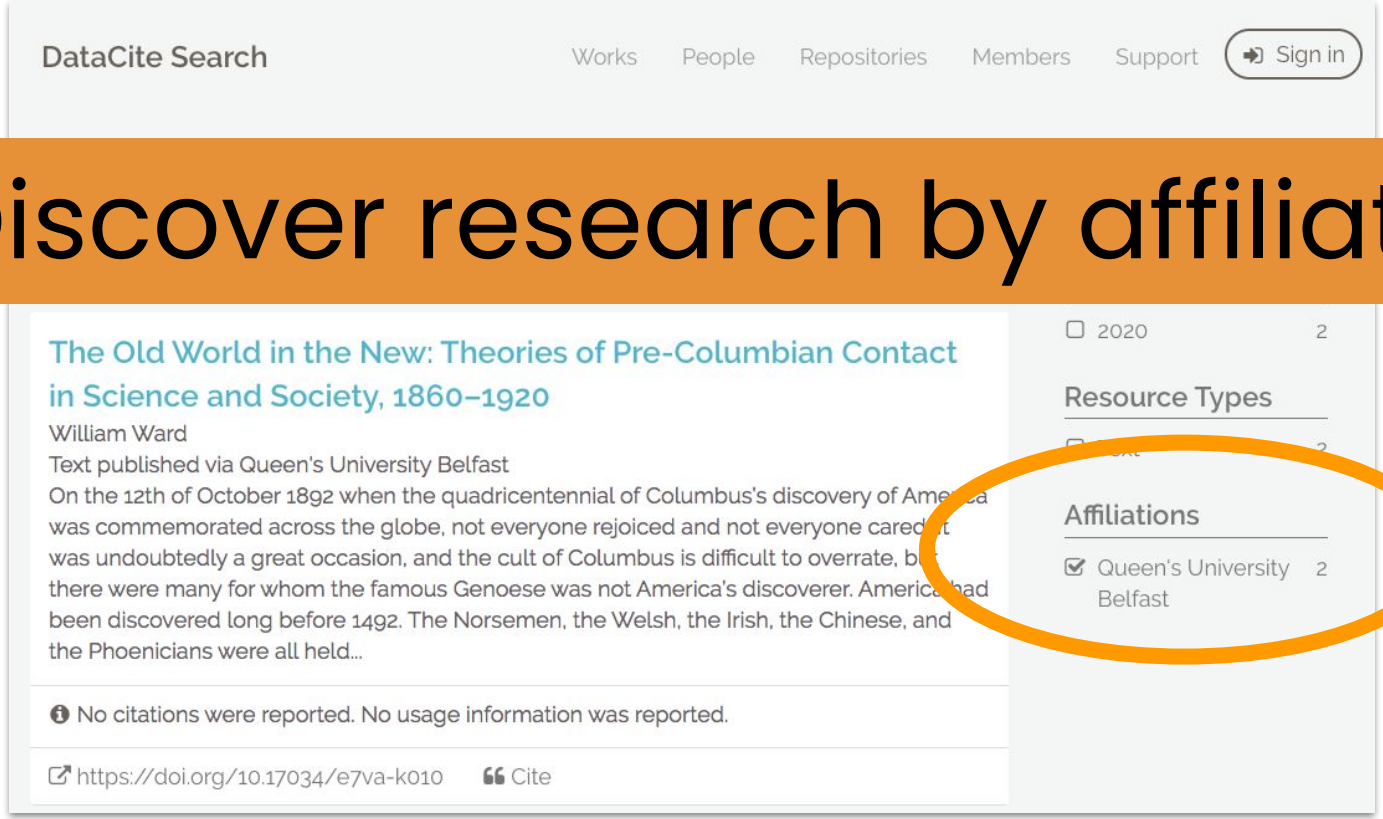
3. Include ROR ID in DOI metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://
datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4/metadata.xsd">
      <familyName>Ward</familyName>
      <nameIdentifier ...>William Ward</
nameIdentifier>
      <affiliation affiliationIdentifier="https://ror.org/00hswnk62"
affiliationIdentifierScheme="ROR" schemeURI="https://ror.org">Queen's University Belfast</
affiliation>
    </creator>
  </creators>
  <titles>
    <title>The Old World in the New: Theories of Pre-Columbian Contact in Science and
Society, 1860–1920</title>
  </titles>
  <publisher>Queen's University Belfast</publisher>
  <publicationYear>2020</publicationYear>
```

ROR

@ResearchOrgs

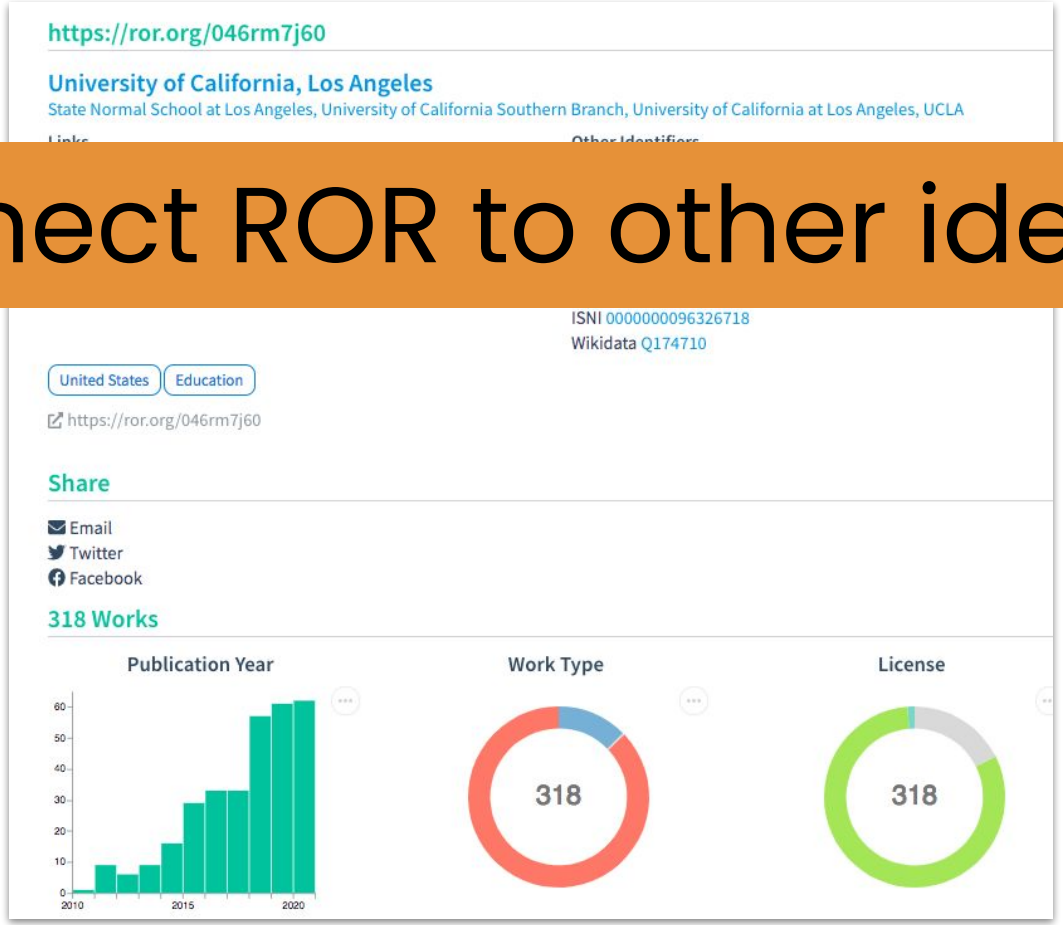
4. Discover research by affiliation



ROR

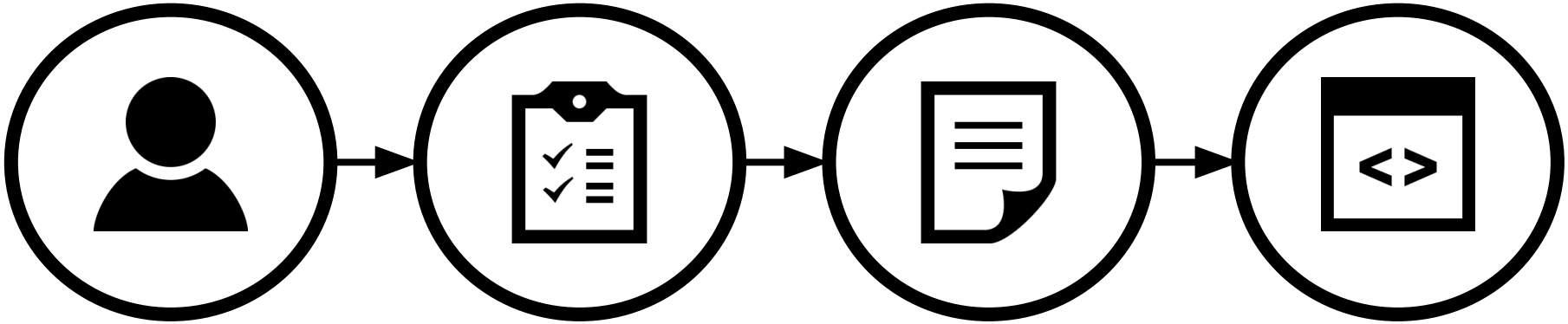
@ResearchOrgs

5. Connect ROR to other identifiers



@ResearchOrgs

Publisher workflow



Researcher selects affiliation from list

Journal checks for competing interests

Paper published with affiliation

ROR ID deposited in Crossref metadata

ROR

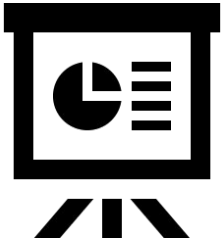
@ResearchOrgs

Further uses of ROR

Journal billing (APC waivers, transformative agreements)

Library licensing negotiations

Search and discovery of research by institution



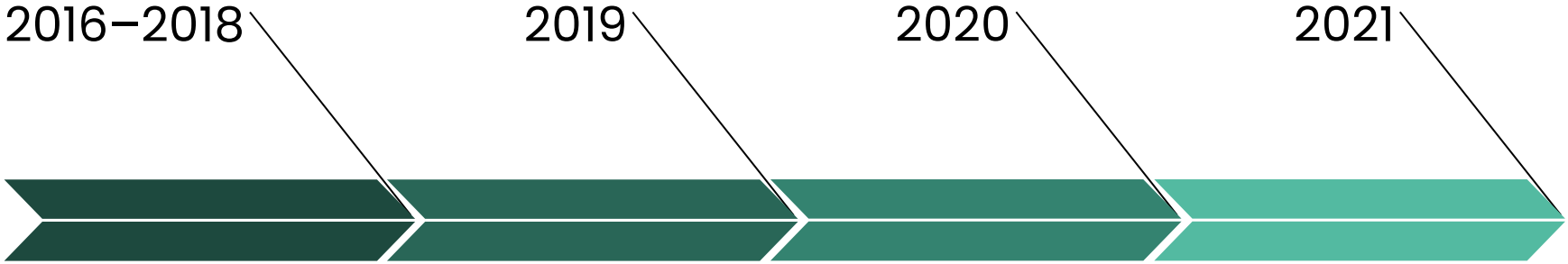
Institutional reporting on research activity

We made it to 2021! A look back and a look ahead

ROR

@ResearchOrgs

Evolution of ROR



Exploratory phase

17 orgs define core vision & basic requirements

Launch of ROR MVR

CDL, Crossref, DataCite, Digital Science launch ROR MVR

Setting up ROR for success

Governance model, resourcing plan, early adopters

Expanding ROR

Strengthen core infrastructure, expand adoption, independent curation

2020 highlights

Recognition

ROR added to USA Library of Congress list of Standard Identifiers

ROR recommended as “standard” persistent identifier for organizations by key stakeholders

ROR commits to following Principles of Open Scholarly Infrastructure (POSI)

Resourcing

2-year Institute for Museum & Library Studies (USA) grant to develop curation model

2-year grant National Science Foundation (USA) grant to drive adoption in research infrastructure

Fundraising through grants and community donations exceeds 2019-2021 goal

ROR team adds technical lead and adoption lead

Adoption

Adoption working group formed

Initial adoption strategy developed

Many integrations emerging in the community

2021 focus areas

Adoption/Integration

- Integration tools, resources, best practices
- API enhancements
- API/adoption metrics & dashboards
- Interoperability with other identifiers & PID services

Infrastructure

- Full GRID metadata support
- Independent updates
- Curation workflows
- Support for inactive records

Sustainability

- Plan for paid service tier
- Long-term governance model
- Community engagement

Adoption & Integration

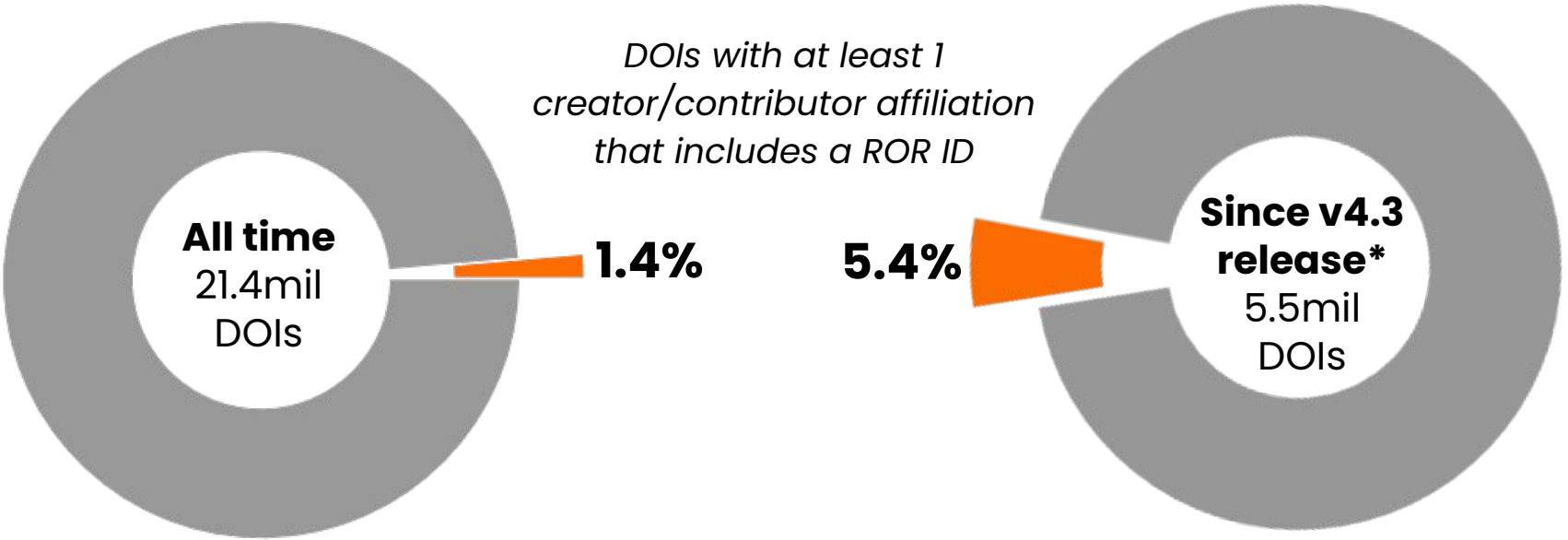
Adoption & integrations: Data downloads

Version	Release Date	Downloads
1.0	Jul 2019	120
2.0	Nov 2019	194
3.0	Feb 2020	36
4.0	Apr 2020	128
5.0	Jul 2020	173
6.0	Oct 2020	112
7.0 (latest)	Dec 2020	16
Total downloads		**779**

- Figshare downloads (to date)
- Repeat customers? Geo data says no
- Data also available in Github (1-2 clones per week)

Adoption & integrations

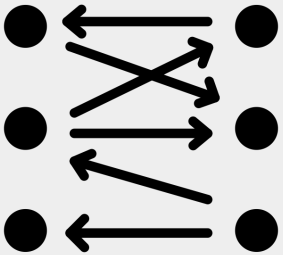
ROR IDs in DataCite DOI metadata



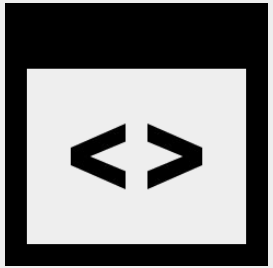
*Organization IDs supported since DataCite schema v4.3, released 16 Aug 2019

Adoption & integrations

Integrations emerging with 3 general functions:



DISAMBIGUATE ORGANIZATIONS
Using ROR API/dataset



PRODUCE METADATA
With ROR IDs included

CONSUME METADATA
With ROR IDs included

Integration highlights

Australian National Data Center (ARDC)

Using ROR to help track investment & outputs across multi partner projects while ensuring correct attribution

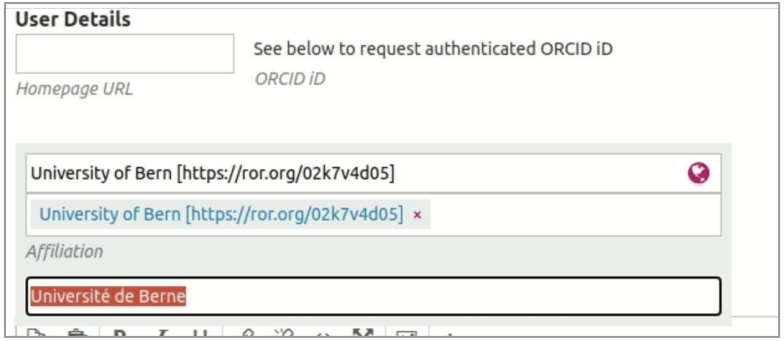
National COVID Cohort Collaborative (N3C)

Using ROR to help manage data transfer/use agreements and IRB info across 500+ participating orgs

cOAlition S Journal Checker Tool

Using ROR to power the institution components of their Plan S compliance checking tool

Integration highlights: OJS



- ROR type-ahead for author affiliations
- Plugin available for OJS v3.2; will be standard in 3.3
- Affiliations not yet included in DOI metadata - planned for June 2021

https://github.com/withanage/ror

Adoption & integrations: Looking ahead

Crossref schema updates with organization ID support coming this year! Planned contributors example:

```
<contributors>
     <person_name sequence="first">
     <given_name>Josiah</given_name>
     <family_name>Carberry</family_na
     <affiliations>
          <institution>
               <institution_name>Brown University</institution_name>
               <institution_id id_type="ror">https://ror.org/05gq02987</institution_id>
               <institution_id id_type="isni">0000000401234567</institution_id>
               <institution_id id_type="wikidata">Q11111111</institution_id>
               <institution_place country="US">Providence, RI</institution_pl e>
               <institution_department>Psychoceramics</institution_depart
          </institution>
     </affiliations>
...
```

Repeatable, to allow for multiple institutions

ROR is the recommended ID type, but others will also be supported.

Adoption & integrations: Looking ahead

Lots planned for 2021!

- API users group
- Adoption/integration metrics & dashboards
- Integration best practices
- More tools/resources to support integration
    - Mapping other ID types to ROR

Adoption & integrations: Looking ahead

We'll need your input - look out for opportunities via our email list/Slack channels.

- Subscribe to ROR emails http://eepurl.com/gjkT9H
- Request to join ROR Slack https://tinyurl.com/ror-slack

Curating ROR

Curation – Background

- Curation has been limited due to dependence on GRID and resourcing constraints
- ROR launched MVR with GRID data. Deliberately did not build curation infrastructure/workflows right away
- Community discussions in 2019 to explore approaches
- 2020 focused on securing resourcing while piloting basic processes

Curation - current status

- Current (temporary) approach:
  - Updates through GRID, on GRID's schedule
  - ROR sends requests to GRID to implement on a limited basis, and ROR syncs GRID's updates
  - Pilot group developing policies & workflows
- New model to begin in 2021
  - ROR no longer dependent on GRID
  - Community-based curation

Vision for ROR curation

- ROR is no longer dependent on GRID for changes
- Anyone can suggest changes to ROR
- Changes are reviewed through centralized and transparent process and with input from community advisors
- Changes are made on a regular and predictable schedule
- Keep as lightweight as possible for sustainability

Curation - Development roadmap

- First step:
  - Basic request & review workflow (similar to now) for independent updates
  - Github issues & pull requests
- Future work:
  - Bulk updates
  - Automated ingestion /ways for integrators to transmit feedback
  - Connections to other identifiers

Let’s ROR together!

https://ror.org
info@ror.org