



# **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



# What is ROR and who is the ROR Community?



# 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>

 <https://ror.org/03yrm5c26>

California Digital Library

CDL

WEBSITE

<http://www.cdlib.org/>

OTHER IDENTIFIERS

GRID [grid.463323.3](#)

ISNI [0000000119575136](#)

Wikidata [Q5020447](#)

UNITED STATES

ARCHIVE



```
← → ↻ 🔒 api.ror.org/organizations
{
  "number_of_results": 97795,
  "time_taken": 9,
  "items": [
    {
      "id": "https://ror.org/01mlpv723",
      "name": "University Hospital of Geneva",
      "acronyms": [
        "HUG"
      ],
      "status": "active",
      "wikipedia_url": "https://en.wikipedia.org/wiki/Geneva_University_Hospitals",
      "labels": [
        {
          "label": "Hôpitaux universitaires de Genève",
          "iso639": "fr"
        }
      ]
    }
  ]
}
```

# ROR API: <https://api.ror.org/organizations>

## ROR Data

321  
views

0  
citations

+ Follow

Version 4 ▾

Published on 03 Apr 2020 - 17:19 by ROR Community

ROR data files.

AUTHORS

ROR Community

CITE THIS COLLECTION

DataCite ▾

CATEGORIES

- Library and Information Studies
- Information Systems

Community, ROR (2019): ROR Data. figshare. Collection.

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

SHARE

community infrastructure

The screenshot shows a web browser window displaying a ROR record. The address bar contains the URL `ror.org/03yrm5c26`, which is circled in orange. An orange arrow points from this URL to the text "ROR ID (URL resolves to record)". Below the browser window, the URL `https://ror.org/03yrm5c26` is also circled in orange. The page content includes the ROR logo, the organization name "California Digital Library" with the acronym "CDL", a "WEBSITE" field with the URL `http://www.cdlib.org/`, and an "OTHER IDENTIFIERS" section listing "GRID grid.463323.3", "ISNI 0000000119575136", and "Wikidata Q5020447". At the bottom, there are two buttons: "UNITED STATES" and "ARCHIVE".

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

The screenshot shows a web browser window with the URL `ror.org/03yrm5c26`. The page features the ROR logo, a search bar, and an 'ABOUT' link. The main content area displays the ROR ID `https://ror.org/03yrm5c26` and the organization name 'California Digital Library' with the acronym 'CDL'. A section titled 'OTHER IDENTIFIERS' is circled in orange and lists the following identifiers: GRID `grid.463323.3`, ISNI `0000000119575136`, and Wikidata `Q5020447`. At the bottom, there are two buttons: 'UNITED STATES' and 'ARCHIVE'.

Crosswalks to other IDs



# ROR reconciler

**OpenRefine** publisher\_affiliation\_names\_with\_local\_ids txt Permalink

Facet / Filter Undo / Redo 8 / 8

Refresh Reset All Remove All

**14 rows** Show as: rows records Show: 5 10 25 50 rows

**name** change 10 choices Sort by: name count Cluster

- All Souls College, University of Oxford 1
- NASA 1
- National Aeronautics and Space Administration 1
- National Institutes of Health 2
- Oxford University 1
- Rinky Dink Foundation 1
- Sloan Foundation 1
- U.S. Department of Transportation 3
- University of Oxford 1

**name: judgment** change 2 choices Sort by: name count

- matched 8
- none 6

	All	publisher_id	name	ror_id
1.	1000001	University of Oxford Choose new match	https://ror.org/052gg0110	
2.	1000002	University of Oxford Choose new match	https://ror.org/052gg0110	
3.	1000003	Wellcome Trust Choose new match	https://ror.org/029chgv08	
4.	1000004	United States Department of Transportation Choose new match	https://ror.org/02xfw2e90	
5.	1000005	National Aeronautics and Space Administration Choose new match	https://ror.org/027ka1x80	
6.	1000006	National Institutes of Health Choose new match	https://ror.org/01cwqze88	
7.	1000007	United States Department of Transportation Choose new match	https://ror.org/02xfw2e90	
8.	1000008	United States Department of Transportation Choose new match	https://ror.org/02xfw2e90	
9.	1000009	National Aeronautics and Space Administration Choose new match	https://ror.org/027ka1x80	
10.	1000010	Wellcome Trust Choose new match	https://ror.org/029chgv08	
11.	1000011	Alfred P. Sloan Foundation Choose new match	https://ror.org/052csg198	
12.	1000012	National Institutes of Health Choose new match	https://ror.org/01cwqze88	

Match this Cell M

**ROR** .community

Location: United States  
ROR ID: https://ror.org/01cwqze88

- Istituto Superiore di Sanità (20)
- National Institutes of Health
- National Health Research Institute
- Instituto Nacional de Salud
- National Institutes of Health
- National Network of Public Health Institutes
- Foundation for the National Institutes of Health
- Ministerul Sanatatii (13)
- Ministero della Salute (12)
- National Institutes of Biomedical Research and Health Services
- Canadian Institutes of Health Research (10)
- National Institutes of Natural Sciences (9)

# Affiliation matching

```
← → ↻ 🔒 api.ror.org/organizations 🔍

{
  "number_of_results": 96793,
  "time_taken": 7,
  "items": [
    {
      "id": "https://ror.org/03b9snr86",
      "name": "Catholic University of Portugal",
      "types": [
        "Education"
      ],
      "links": [
        "http://www.ucp.pt/site/custom/template/ucptplportalhome.asp?sspageid=1&lang=1"
      ],
      "aliases": [],
    }
  ]
}
```

# 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 <a href="#">policies</a>
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 <a href="#">policies</a> for more details
external_ids	External ID	False	Other IDs known to refer to the institute. See our <a href="#">policies</a> 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	Latitude 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

id	sho...	long_name	abb...	created_at	updated_at	ror_id
84	NULL	VU University Amsterdam	N...	2019-07-30 22:16:21	2019-07-30 22:16:21	<a href="https://ror.org/008xzew50">https://ror.org/008xzew50</a>
85	NULL	China University of Geosciences	N...	2019-07-30 22:16:21	2019-11-14 00:52:12	<a href="https://ror.org/04gcegc37">https://ror.org/04gcegc37</a>
86	NULL	Technical University of Zvolen	N...	2019-07-30 22:16:22	2019-07-30 22:16:22	<a href="https://ror.org/00j75pt62">https://ror.org/00j75pt62</a>

## 2. Store ROR ID

id	sho...	long_name	abb...	created_at	updated_at	ror_id
95	NULL	Czech University of Life Sciences Prague	N...	2019-07-30 22:16:28	2019-07-30 22:16:28	<a href="https://ror.org/0415vcw02">https://ror.org/0415vcw02</a>
96	NULL	VNU University of Science	N...	2019-07-30 22:16:30	2019-07-30 22:16:30	<a href="https://ror.org/05w54hk79">https://ror.org/05w54hk79</a>
97	NULL	York University	N...	2019-07-30 22:16:30	2019-07-30 22:16:30	<a href="https://ror.org/05fq50484">https://ror.org/05fq50484</a>
98	NULL	Istanbul University	N...	2019-07-30 22:16:31	2019-07-30 22:16:31	<a href="https://ror.org/03a5qrr21">https://ror.org/03a5qrr21</a>
99	NULL	Kagoshima University	N...	2019-07-30 22:16:31	2019-07-30 22:16:31	<a href="https://ror.org/03ss88z23">https://ror.org/03ss88z23</a>
100	NULL	University of New Brunswick	N...	2019-07-30 22:16:32	2019-07-30 22:16:32	<a href="https://ror.org/05nkf0n29">https://ror.org/05nkf0n29</a>
101	NULL	Gadjah Mada University	N...	2019-07-30 22:16:33	2019-07-30 22:16:33	<a href="https://ror.org/03ke6d638">https://ror.org/03ke6d638</a>
102	NULL	Karadeniz Technical University	N...	2019-07-30 22:16:33	2019-07-30 22:16:33	<a href="https://ror.org/03z8fy40">https://ror.org/03z8fy40</a>
103	NULL	University of British Columbia	N...	2019-07-30 22:16:33	2019-07-30 22:16:33	<a href="https://ror.org/03rmrcq20">https://ror.org/03rmrcq20</a>
104	NULL	University of Toronto	N...	2019-07-30 22:16:34	2019-07-30 22:16:34	<a href="https://ror.org/03dbr7087">https://ror.org/03dbr7087</a>
105	NULL	University of Twente	N...	2019-07-30 22:16:35	2019-07-30 22:16:35	<a href="https://ror.org/006hf6230">https://ror.org/006hf6230</a>
106	NULL	Philipp University of Marburg	N...	2019-07-30 22:16:35	2019-07-30 22:16:35	<a href="https://ror.org/01rdrb571">https://ror.org/01rdrb571</a>
107	NULL	University of Lausanne	N...	2019-07-30 22:16:35	2019-07-30 22:16:35	<a href="https://ror.org/019whta54">https://ror.org/019whta54</a>
108	NULL	Utrecht University	N...	2019-07-30 22:16:35	2019-07-30 22:16:35	<a href="https://ror.org/04pp8hn57">https://ror.org/04pp8hn57</a>
109	NULL	Waseda University	N...	2019-07-30 22:16:35	2019-07-30 22:16:35	<a href="https://ror.org/00ntfnx83">https://ror.org/00ntfnx83</a>
110	NULL	Monash University	N...	2019-07-30 22:16:36	2019-07-30 22:16:36	<a href="https://ror.org/02bfwt286">https://ror.org/02bfwt286</a>
111	NULL	Shimane University	N...	2019-07-30 22:16:37	2019-07-30 22:16:37	<a href="https://ror.org/01jaaym28">https://ror.org/01jaaym28</a>
112	NULL	University of Iceland	N...	2019-07-30 22:16:37	2019-07-30 22:16:37	<a href="https://ror.org/01db6h964">https://ror.org/01db6h964</a>
113	NULL	University of Tsukuba	N...	2019-07-30 22:16:38	2019-07-30 22:16:38	<a href="https://ror.org/02956yf07">https://ror.org/02956yf07</a>
114	NULL	Manchester University	N...	2019-07-30 22:16:39	2019-07-30 22:16:39	<a href="https://ror.org/02aswte26">https://ror.org/02aswte26</a>
115	NULL	University of Manchester	N...	2019-07-30 22:16:39	2019-07-30 22:16:39	<a href="https://ror.org/027m9bs27">https://ror.org/027m9bs27</a>
116	NULL	Fukui Prefectural University	N...	2019-07-30 22:16:42	2019-07-30 22:16:42	<a href="https://ror.org/02c3vg160">https://ror.org/02c3vg160</a>
117	NULL	Kiel University	N...	2019-07-30 22:16:42	2019-07-30 22:16:42	<a href="https://ror.org/04v76ef78">https://ror.org/04v76ef78</a>

```
<?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">
```

## 3. Include ROR ID in DOI metadata

```
  <givenName>William</givenName>
  <familyName>Ward</familyName>
  <nameIdentifier nameIdentifierScheme="Other" schemeURI="http://orcid.org/0000-0001-9142-1000">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>
```

# 4. Discover research by affiliation

## The Old World in the New: Theories of Pre-Columbian Contact in Science and Society, 1860–1920

William Ward

Text published via Queen's University Belfast

On the 12th of October 1892 when the quadricentennial of Columbus's discovery of America was commemorated across the globe, not everyone rejoiced and not everyone cared. It was undoubtedly a great occasion, and the cult of Columbus is difficult to overrate, but there were many for whom the famous Genoese was not America's discoverer. America had been discovered long before 1492. The Norsemen, the Welsh, the Irish, the Chinese, and the Phoenicians were all held...

**i** No citations were reported. No usage information was reported.

<https://doi.org/10.17034/e7va-k010> **“** Cite

2020 2

### Resource Types

2

### Affiliations

Queen's University Belfast 2

<https://ror.org/046rm7j60>

## University of California, Los Angeles

State Normal School at Los Angeles, University of California Southern Branch, University of California at Los Angeles, UCLA

Links

Other Identifiers

# 5. Connect ROR to other identifiers

ISNI 0000000096326718

Wikidata Q174710

United States Education

<https://ror.org/046rm7j60>

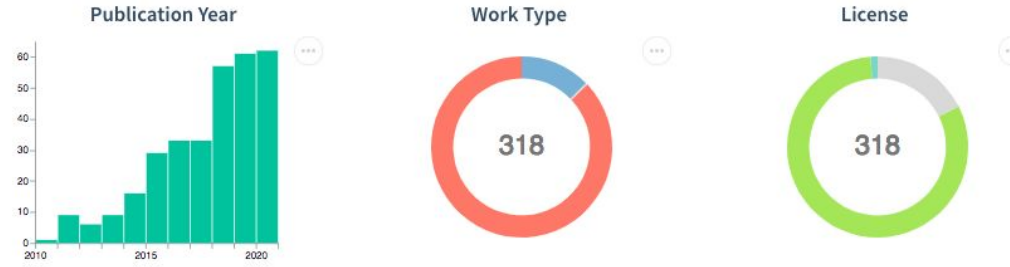
### Share

Email

Twitter

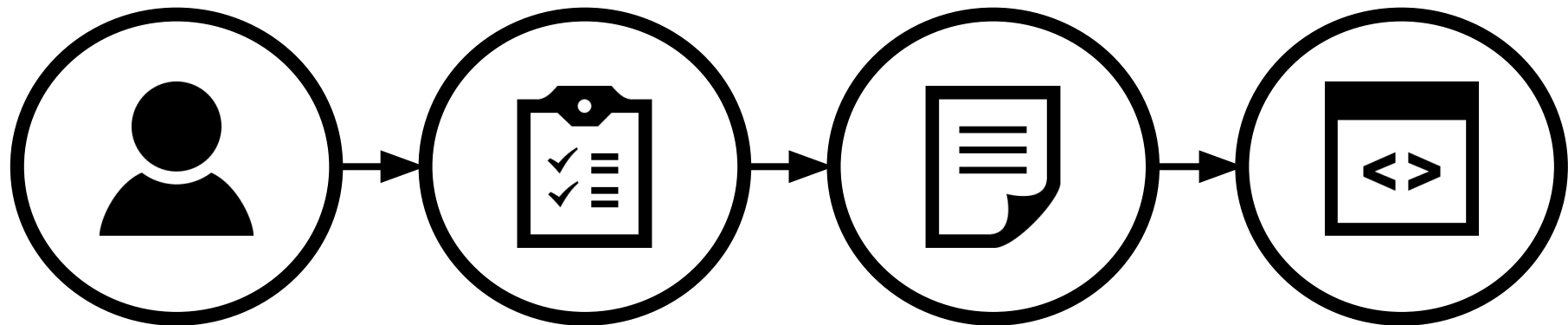
Facebook

### 318 Works





# Publisher workflow



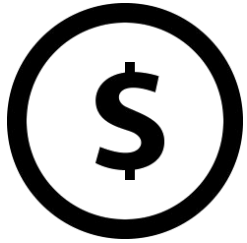
Researcher  
selects  
affiliation  
from list

Journal  
checks for  
competing  
interests

Paper  
published  
with  
affiliation

ROR ID  
deposited  
in Crossref  
metadata

# Further uses of ROR



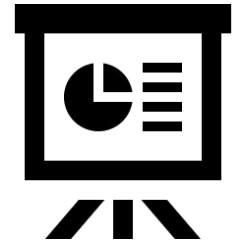
Journal billing  
(APC waivers,  
transformative  
agreements)



Library licensing  
negotiations



Search and  
discovery of  
research by  
institution

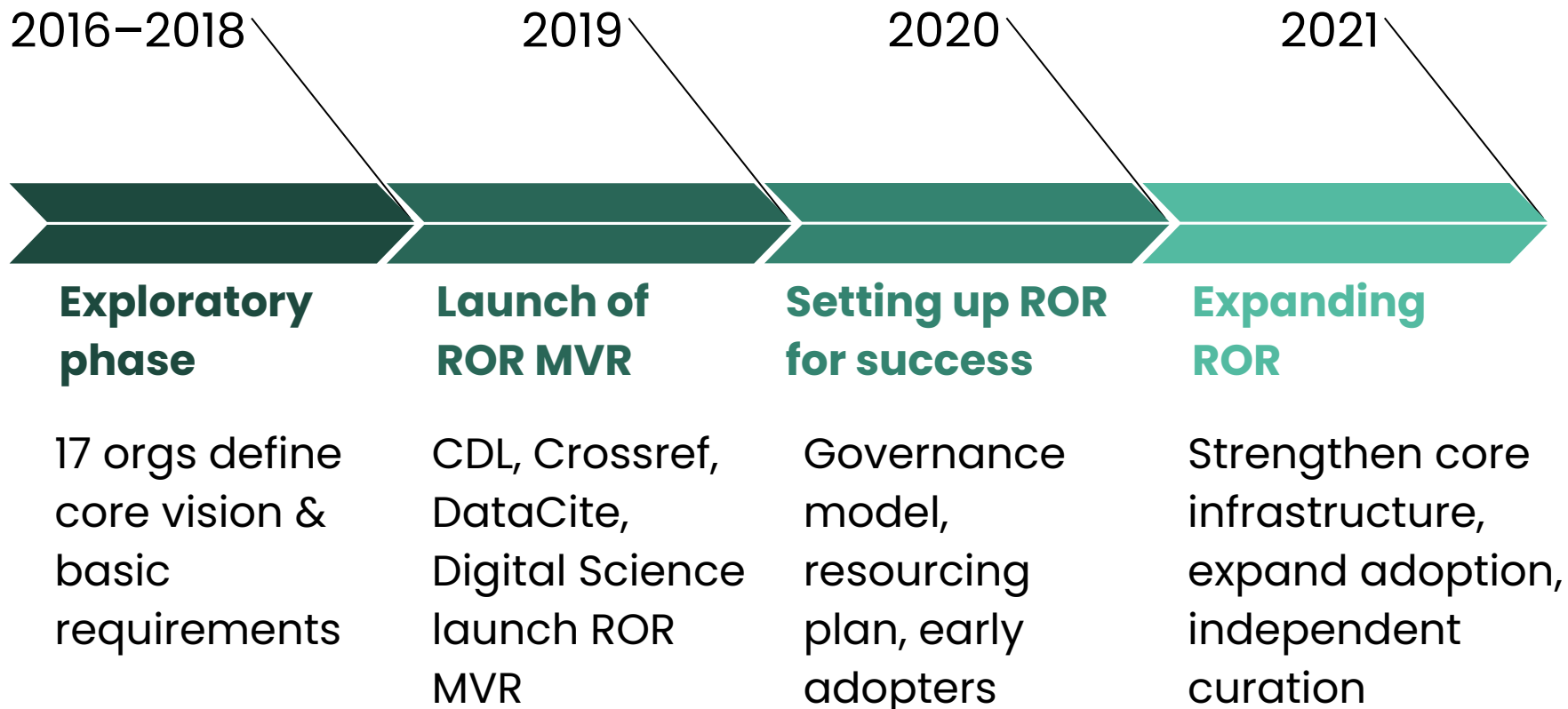


Institutional  
reporting on  
research  
activity



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

# Evolution of ROR



# 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

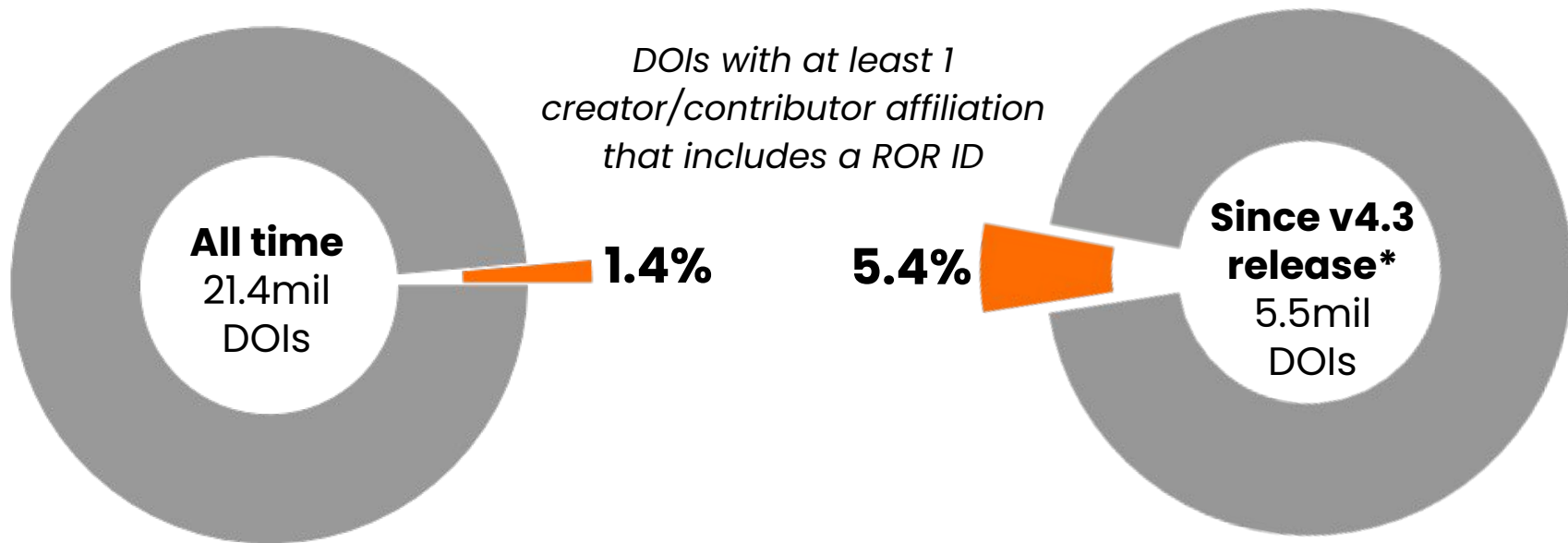
# Adoption & integrations: Data downloads

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
<b>Total downloads</b>		<b>779</b>

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

# Adoption & integrations

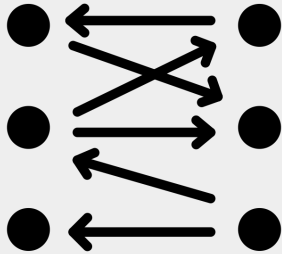
## ROR IDs in DataCite DOI metadata



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

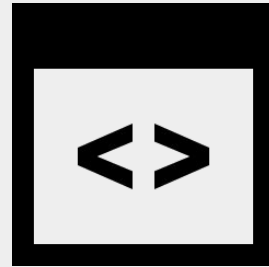
# Adoption & integrations

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

**User Details**

See below to request authenticated ORCID iD  
*Homepage URL* *ORCID iD*

- University of Bern
- University of Pisa
- University of Basel

**User Details**

See below to request authenticated ORCID iD  
*Homepage URL* *ORCID iD*

*Affiliation*

- 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_name>
  <affiliations>
    <institution>
      <institution_name>Brown University</institution_name>
      <institution_id id_type="ror">https://ror.org/05gg02987</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_place>
      <institution_department>Psychoceramics</institution_department>
    </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](mailto:info@ror.org)

