Curating a Community Registry of Research Organizations

California Digital Library (CDL) requests $247,415 for a National Leadership Grant for Libraries (NLG-L) Project Grant in the National Digital Infrastructures and Initiatives category to develop and implement a curation model for the Research Organization Registry (ROR). ROR has begun piloting a community-based curation model in its first year of operation and now seeks funding to expand and sustain this approach as the registry matures.

Statement of National Need

The Research Organization Registry (ROR) launched in January 2019[1] with a mission to provide open, trusted, and noncommercial infrastructure for research organization IDs and associated metadata. ROR is a comprehensive and authoritative registry that comprises close to 100,000 organizations (as of March 2020), each with a unique ID. The registry is managed as a partnership by California Digital Library, Crossref, and DataCite, following a collaborative planning process by seventeen different organizations[2] to define requirements for a community-owned registry of organization IDs and to select a core steering group to bring this registry to fruition.

ROR emerged to fill a specific and crucial gap in scholarly infrastructure: how to effectively and unambiguously identify the organizations that are affiliated with research outputs. Persistent identifiers (PIDs) are becoming increasingly more central to the research landscape in the United States[3] and beyond, facilitating access to research as well as the tracking of research use and impact. ORCID IDs make it possible to identify and disambiguate researchers, and digital object identifiers (DOIs) for articles, datasets, dissertations, and other outputs are crucial for discovery and citation. But how do we know which *institutions* are associated with this research? Until ROR emerged, there was no easy and open way to answer this question.

Identifying all research outputs from a given research organization should be simple, but this vital information can be surprisingly difficult to obtain, either because affiliation data is collected in proprietary commercial systems and locked behind paywalls, or because it may exist only as unstructured text ("University of California-Los Angeles" vs. "UCLA"). In some cases, affiliation information may not be collected at all. The need for data on publications and research outputs by institution by librarians and academic administrators is increasing for a variety of reasons, whether to report on scholarly output and impact, track research supported by a specific funder or in compliance with a federal policy, establish a baseline for transformative agreements with journals to support Open Access publications, determine which journals are publishing research by their institution, and others. The lack of easy and open access to this data puts institutions and librarians in the position of

---

[1] https://ror.org/blog/2019-02-10-announcing-first-ror-prototype/
[2] https://ror.org/about/#history
[3] National Science Foundation, "Dear Colleague Letter: Effective Practices for Data," May 20, 2019. https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

having to pay and rely on commercial vendors for access to their own information, or having to undergo inefficient and duplicated efforts to analyze and reconcile multiple data sources.

ROR addresses these challenges by providing a registry of standardized, globally unique identifiers that are freely and openly available under a CC0 license and can be used for the structured and unambiguous collection of researcher affiliations. ROR also provides a public API and additional tools for querying, manipulating, and integrating its dataset. DataCite and Crossref, which collectively register DOIs for the majority of scholarly outputs globally, are supporting ROR IDs in their metadata, which enables clean and powerful affiliation data for every object with a DOI. ROR IDs can be integrated wherever affiliations are collected (e.g., journal submission platforms, data repositories, funder portals, institutional repositories). Scholarly information systems for tracking grants, data management plans, research impact, and more can benefit from ROR's open infrastructure.

Research organizations are not static entities. New organizations are always forming and existing organizations regularly change names, update their websites, merge with or split from other organizations, or shut down entirely. These are inevitable changes that ROR must have the capacity to respond to and reflect if its dataset is to be a comprehensive, useful, and trusted source of information, and if its infrastructure is to be implemented widely in scholarly information systems. We believe that ROR can accomplish these goals by building a curation model for the registry, focusing on three key milestones: (1) establish a community-based curation advisory board; (2) support a curation coordinator; and (3) complete the technical work required to enable curation tasks. These three components will result in setting up ROR to handle curation at scale and to support responsible stewardship of registry data both on a day-to-day basis and over the long term.

IMLS funding to develop and implement this approach to curation will allow ROR to successfully manage and curate its data in ways that reflect the changing nature of research organizations, respond to community needs, and reinforce librarian values around trusted information sources and systems. It will allow ROR to make its data and infrastructure available to stakeholders who need and can benefit from it, regardless of their ability to pay for or access it.

While we have already begun piloting efforts toward these milestones, as outlined below, IMLS funding will allow us to accelerate this work by providing much-needed resources that supplement those provided in kind by ROR's governing organizations. The sooner ROR has a high-functioning, scalable curation model, the more successfully ROR's infrastructure can be used and adopted.

Project Design

Curation advisory board
We have already been piloting community-based curation processes on a small scale with three members of the ROR Community Advisory Group, who hold positions at a US academic library,

government science agency, and publisher. We plan to expand this pilot group and establish an official curation advisory board, inaugurating the first cohort in Fall 2020. The curation advisory board will provide expert input to guide decision-making about changes to the registry and the long-term management of registry data. This input will include:

- Refining and reinforcing the criteria used to determine when new organizations should (or should not) be added to the registry, in line with ROR's stated scope

- How to determine authoritative sources of information about institutional metadata (i.e., the official name or acronym of a university)

- How to ensure that ROR serves a global and multilingual audience with appropriate metadata fields for names in multiple languages and with multilingual support for API queries and front-end search functionality

- How to best capture changes to records over time (i.e., what details to log when a record is changed, and how records might be versioned to capture changes over time)

- How the ROR data model might be adjusted to meet community needs (i.e., adding new metadata fields according to community priorities)

- Which workflows and functionality should be put in place to triage and coordinate requests in an efficient, scalable way and in line with community needs

The curation advisory board will be modeled after journal editorial boards and grant proposal study sections. Board members will bring specific expertise to ROR so that the registry's curation model can effectively embody the heterogeneity of research organizations within a controlled framework for managing organizational metadata. We will seek advisory board members who are experienced in non-English languages, non-Western institutions, and scholarly communications in the Global South, and who represent different institution types and geographic regions, particularly underrepresented areas within the United States. A breadth of expertise from advisory board members will help ROR to navigate decisions around how institutions should be represented in the registry, how ROR metadata should support effective management of institutional records, and how to ensure registry data is globally accessible, discoverable, and useful.

We will solicit board nominations from groups at all career stages and across the library, scholarly communication, digital curation, open infrastructure, and PID communities. Existing members of ROR's broader community advisory group[4] will be involved in the nomination and selection process. This group has already engaged in discussions and preliminary planning about curation policies and processes, confirming that there is an existing base of community experts willing to actively contribute to ROR, and reinforcing the importance of involving institutions, librarians, metadata

---

[4] https://ror.org/supporters/#ror-community-advisors

experts, and research administrators in the registry's curation. We will also communicate opportunities for participation through listserv announcements, social media, and blog posts. Prospective members interested in joining the board will be asked to submit a simple questionnaire about their reasons for wanting to be involved and the type of expertise they would like to contribute.

Board participation is an opportunity to have a say in developing community-led infrastructure. It is also an opportunity for members—especially those early in their career stages—to raise their profiles with service to a global project that can be listed on CVs, ORCID records, and elsewhere. Board members will be publicly recognized on the ROR website and in all related communications.

Advisory board cohorts will serve for a term of one year with an opportunity to extend. We will seek cohorts of 8-10 members depending on the nomination pool and on specific expertise required for ROR's curation needs. Most of the board's interaction will take place asynchronously via email, Google Docs, and other collaboration tools. There will be one conference call per month to discuss any major decisions that need to be made. The core tasks of reviewing proposed changes and/or policy and workflow updates and participating in a monthly decision call will represent 1-2 hours per month. Board members may take on additional tasks and projects according to their availability and interests.

Given the importance of developing appropriate and effective policies drawing on the broad expertise of curation board members, we anticipate holding one face-to-face meeting per year for in-depth discussions and extended collaborative work sessions. Meetings will be collocated with major conferences to facilitate travel logistics, and we will provide travel funding to ensure all that all board members are able to attend regardless of their financial situations.

We believe that establishing a curation advisory board is a scalable and effective way to involve community stakeholders in ROR's curation and to help shape ROR's growth and development. Given that we are already piloting this model with existing community members, and given the strong enthusiasm for ROR's community-based ethos that we have seen through outreach events, we believe that there will be support for this model and that we can draw on an existing base of contributors who are already invested in ROR's success. We have experience coordinating community-based projects through other activities at the University of California, including the collaborative development model employed by the DMPTool[5] as well as CDL's leadership in community-owned data publishing initiatives such as Dryad[6] and in facilitating conversation and support for open-source data publishing systems (as in the IMLS grant project LG-73-18-0196-18). We believe that adopting a similar model for ROR is a feasible approach and can be done in a meaningful and efficient way.

---

[5] https://blog.dmptool.org/2018/02/27/new-dmptool-launched-today/
[6] https://www.eurekalert.org/pub_releases/2018-05/cdl-adp052918.php

Curation coordinator

Curation-related tasks are currently being performed by the ROR project lead on a limited basis. Funding from IMLS will allow ROR to support a dedicated staff person to ensure that curation tasks are coordinated and completed efficiently and that curation operations can scale. The coordinator will, as now, work closely with the curation advisory board and ROR technical lead.

The curation coordinator will be the first point of contact for requests to add or update ROR records. ROR currently receives approximately thirteen requests per month on average, but this volume is likely to increase as awareness of ROR grows and as ROR IDs become more widely adopted in scholarly infrastructure. For example, requests spiked when Dryad implemented an affiliation field connected to ROR, as some researchers submitting datasets did not find their affiliations in ROR's database and contacted ROR to request that a new record be added. National registries of institutions are also interested in providing bulk data donations to ROR, which will also lead to increased request volumes when ROR's curation infrastructure has the capacity to handle mass ingests. Requests will be triaged and prioritized by the curation coordinator by request type and priority. The curation coordinator will follow up with requestors for more information as needed, make recommendations for action, prepare decisions for input from the advisory board, and coordinate the technical changes and quality assurance to ensure the update is completed successfully in ROR's database.

The curation coordinator will develop and refine the workflows by which requests are submitted to ROR, how these requests are triaged and reviewed, and how the resulting curation decisions are made and communicated. We plan to use Google Forms/Google Docs and Github issues for the initial workflows, although other tools and processes may be introduced later on to maximize efficiency and transparency. The coordinator will also work with the ROR development lead to define technical requirements for curation.

The curation coordinator will ensure that policy decisions are documented and that registry updates are communicated in public release notes. Request volume will be tracked over time and the coordinator will share reports and data on a regular basis with ROR project leadership and the broader community.

We will conduct outreach among the ROR community and other stakeholders to solicit ideas and suggestions for the full scope of this role. We plan to have the position duties fully and appropriately defined and outlined by the time the grant period begins in September 2020.

Technical development

ROR was launched as a minimum viable product with seed data provided by Digital Science's GRID dataset of institutions, and the registry remains dependent on GRID for curation as ROR has lacked

adequate development resources to update the registry's infrastructure to support independent curation. A key development milestone for ROR, and a key priority for ROR's stakeholders, is functionality to make changes in ROR independently without coordinating these changes through GRID. This will allow ROR to fully implement its own curation policies and to release regular registry updates and data dumps on its own schedule and according to community needs. The pace of this development work has been limited thus far given that ROR has only had ad hoc developer support contributed in kind by its partner organizations. In early 2020, we brought on a dedicated part-time technical lead donated by Crossref to supplement ad hoc development support, but we still need to accelerate the technical work to encourage greater trust and adoption as quickly as possible.

ROR's technical team has drafted preliminary requirements for a basic curation approach that we will be starting to implement around the middle of 2020. We will be taking on more complex development work toward the end of 2020 when the grant period begins. Subsequent development priorities include determining how to version records over time, how to add additional metadata fields, and how to ingest bulk data donations from national or regional registries. Ultimately, we also plan to develop an administrative layer for curation tasks to support functionality such as specific roles and permissions as well as review and approval workflows.

Target community and key collaborators
Establishing a curation model for ROR is a crucial step in reinforcing the registry's role within the library and scholarly research communities by ensuring the long-term quality and stewardship of its data. This is a natural extension of previous groundwork to develop the registry as a community-led infrastructure project. The initial collaborations that resulted in ROR involved a broad cross-section of organizations working in the scholarly communication landscape. ROR is currently led by three organizations—California Digital Library, Crossref, and DataCite—with deep expertise in research, digital libraries, publishing, data curation, and open infrastructure. ROR is also guided by a steering group[7] of representatives from trusted organizations in libraries, networked research, and research infrastructure, both nationally and globally, and a community advisory group of key stakeholders located around the world. ROR project team members are involved in and have presented to national and global library and scholarly communication communities, including FORCE11, OpenCon, Library Carpentry, Coalition for Networked Information, and the Research Data Alliance.

The ROR curation advisory board creates an opening for librarians and other research stakeholders to become more involved in the development and management of open scholarly infrastructure projects focused on leveraging persistent identifiers to increase public access. Academic librarians are already at the forefront of enabling PIDs on their campuses, whether encouraging researchers to obtain ORCIDs or helping to register DOIs for datasets and publications. Academic librarians are also fulfilling key roles in supporting the tracking of research outputs for their institutions, whether for

---

[7] https://ror.org/about/#steering-group

Open Access mandates, funder policies, or administrative reports. As PIDs become more embedded in national conversations, it is crucial for librarians to be involved. Librarian expertise on the advisory board will be crucial for ROR to be a trusted source of information and to be seen, valued, and used as a community good.

Project development

Upon launching the registry in January 2019, the ROR team began scoping different approaches to the registry's curation by exploring the technical requirements needed to support curation as well as policies and procedures to guide effective management of ROR data over time. This exploratory work involved stakeholder input through regular calls with ROR's community advisory group as well as through presentations and webinars. Following this exploratory period, ROR began a pilot approach to curation in Fall 2019 with an inaugural cohort of advisors from the community advisory group. This cohort has already begun to review requests for changes and additions to the registry and to provide input on draft curation policies, working with the ROR project lead. In early 2020, ROR brought on a dedicated part-time technical lead to begin updating ROR's infrastructure to enable curation.

Project funding from IMLS will allow us to move beyond this pilot phase. We are ready to scale curation workflows and accelerate development work, but we need funding to secure dedicated staffing (for curation coordination and technical development) and expand the curation advisory board. By the time the grant period begins, we will have a curation coordinator role in place and we will be increasing developer time on the project. We will announce a call for advisory board nominations in Fall 2020. By the end of the grant period, we expect to have reached a maintenance stage for the project whereby the workflows and advisory board activities are stable, predictable, and sustainable over the long term.

Project evaluation and outcomes

Since ROR launched, we have been tracking the volume and types of incoming requests for changes to the registry. This has provided a benchmark for estimating the volume and nature of requests over time and anticipating the technical functionality that will be needed to support curation. One key outcome of curation board activities during the grant period will be collecting and sharing comprehensive data on the specific quantity and nature of curation requests, and how these requests were resolved. This data will be valuable for ROR's projections into the future and will help us to develop more standardized workflows for evaluating requests and making decisions. It will also be valuable for similar community projects that may benefit from having a baseline of what to expect. We will ensure that our documentation is publicly available (via Github, Google Docs, and the ROR website) and we will share data and details of our approach in blog posts, articles, and presentations.

As a community-driven effort, ROR relies on community feedback to guide its growth and measure its progress. We will conduct surveys with ROR community users and solicit input from the broader community advisory group as we develop and implement the curation model. We will also schedule

quarterly check-ins with the curation advisory board to collect feedback and make necessary adjustments. Survey instruments and calls for feedback will be focused on evaluating the effectiveness of a centralized community-based approach as well as the quality of the guidelines, communications, and results that are shared with the community.

Throughout the grant period, we will measure and recalibrate resourcing needs to determine how ROR's curation model might evolve. By the end of the grant period, we expect to have a predictable and sustainable model that can be carried forward.

Assumptions

The project of establishing this approach to ROR's curation is motivated by several assumptions:

First, ROR's community-based approach to curation has emerged in response to feedback from ROR stakeholders. Early concepts for ROR during initial planning stages included possible options for institutions—via designated authorized representatives—to manage and update their own records when needed. However, since the registry's official launch in 2019, discussions with ROR stakeholders have not revealed an explicit or overwhelming interest in this functionality. Rather, stakeholders are in support of a more centralized approach in which a group of community members is trusted to make curation decisions.

Second, centralizing the registry's curation is more efficient and scalable than identifying a designated representative at each of the nearly 100,000 organizations included in the registry and building additional infrastructure on top of ROR to manage these representatives' access to the registry through logins and permissions. A centralized board can apply curation approaches at scale to ensure greater consistency and quality of registry data—especially concerning practices and decisions that affect multiple records. Coordinating registry curation through a group of advisors will also help to establish clear and straightforward procedures for receiving, reviewing, and acting upon requests. Currently, requests arrive via email to a general ROR inbox where they are triaged by the ROR project lead. With a centralized curation model, requests can be received and tracked centrally using Google Forms and Google Docs and then converted to issues on a Github board, where resulting registry updates can be communicated transparently in public release notes.

Lastly, we believe that the work of maintaining the ROR registry will be best accomplished by leveraging a curation advisory board that includes representatives of academic institutions, especially librarians, who have the expertise and the ethos to help develop appropriate curation policies and processes to ensure the integrity and longevity of registry data in line with community-based values. For this reason, we see this curation model as a key mechanism to ensure that librarians are included in this project.

Risks

Given that we have already been monitoring the scope of curation requests for ROR and piloting preliminary community-based curation workflows, we feel confident that the proposed approach will be successful and believe that we are prepared to mitigate any potential risks in expanding this model. One concern that may arise is about advisory board members' time and labor. To anticipate and offset these concerns, we will provide a clear summary of the scope and responsibilities of board members and estimates of how much time their contributions will entail. The curation coordinator will also be able to manage and constrain the time needed from board members by prioritizing tasks and decisions. We will officially recognize board involvement as professional service so that members can rely on this as justification to perform duties during regular work hours. If board members are eager to participate but concerned about time commitments—either up front or during the course of their tenure—we will adjust tasks and requests for input accordingly so they can remain involved.

It is also crucial that ROR be able to maintain and apply a consistent set of curation criteria that is supported and deployed by curation board members, especially as board membership changes over time. We will mitigate this risk with comprehensive documentation and robust onboarding procedures for new board members so that they have a common basis for decision-making.

National Impact

ROR is uniquely focused on accessible infrastructure by and for the scholarly community. The overarching ROR initiative stands to impact both national and global research stakeholders in a number of ways. Open infrastructure for institutions to identify and track their research outputs gives librarians and administrators a unique opportunity to control and leverage their own data and to avoid relying on commercial vendors for this information. This data is immensely and increasingly valuable for institutions on multiple levels, whether for tracking and capturing key insights such as which of their researchers' outputs have been supported by a particular funder, which research outputs are available Open Access, or which of their publications might be affected by a high-profile license negotiation as in the University of California's recent cancellation of its contract with Elsevier,[8] resulting in the UC Libraries' loss of access to a substantial amount of Elsevier journal content, which includes UC-produced research. ROR is the missing piece of the persistent identifier puzzle that can now join ORCIDs for researchers and DOIs for research outputs to complete the picture and leverage the power of PIDs to discover, connect, and analyze research.

ROR represents a unique initiative because it is the first and only effort to build open infrastructure for organization IDs that is both specifically focused on identifying academic affiliations *and* deliberately designed with community input. The registry's launch in January 2019 was met with great interest and enthusiasm, and the registry's first year saw a number of early implementations of ROR IDs in various systems. This amount of support and uptake so early in the project is a strong indicator of ROR's longer-term impact. However, while the registry already exists and ROR IDs are

---

[8] https://osc.universityofcalifornia.edu/2019/03/open-statement-why-uc-terminated-journal-negotiations-with-elsevier/

already in use, the project's next phase needs to be focused on how best to grow and maintain the registry. A community-based curation model for ROR is essential for ROR's success and potential impact by ensuring that library values are embedded in and guiding the way in which ROR manages its data. Too often, librarians are excluded from scholarly infrastructure projects and are not given a say in the design and implementation of the systems on which they rely. The curation advisory board will enable ROR to support effective curation of the registry and express the project's community-based ethos through a centralized framework.

IMLS funding to support ROR's curation activities is vital at this stage of the project. In ROR's first year, the project has been sustained by in-kind contributions from its steering organizations (as set forth in agreements between these organizations). ROR launched a sustainability campaign in October 2019 focused on generating startup funding through community investments and grants to develop and establish the registry's infrastructure and curation workflows. ROR plans to introduce an optional paid service tier in 2022 to cover the basic costs of maintaining the registry while keeping the registry's core data open and free. At that time, it will transition away from relying on external grants and contributions.

We believe that we can keep the costs of running the registry relatively low. We are keeping the scope of the effort contained by focusing on infrastructure for research organizations and we are keeping the overhead minimal by developing this infrastructure as a partnership that does not require establishing a new legal entity to run it. Basic registry costs include technical logistics such as server hosting, along with other expenses to support communications and community engagement (i.e., hosting annual community meetings), and staffing of core FTEs specializing in technical development, project coordination, user support/adoption, and curation. ROR's steering organizations will also continue to contribute in-kind resources to the project. By launching a paid service tier in 2022 focused on covering costs and supporting staffing, we will be able to sustain ROR curation over the long term.

The approach to curation that we are proposing has the potential for even broader impact by serving as a model to be adapted by other community-based infrastructure projects. Developing open community infrastructure presents many challenges. Projects can struggle to gain legitimacy within specific communities, or fail to create specific roles and mechanisms for community members to contribute. Projects also face financial challenges: providing infrastructure for free still requires funds and resources, which may not be readily available or perceived to be necessary. ROR's approach to community-driven curation has the potential to pilot a unique model for driving and sustaining community engagement in a meaningful and effective way. It is also an opportunity to support long-term sustainability of open community infrastructure without creating extensive overhead or barriers to participation. Connecting the library and PID communities through ROR's curation model will leverage the unique expertise of each and reinscribe the importance of librarian values to open infrastructure initiatives.

Schedule of Completion: Curating a Community Registry of Research Organizations

Task	Sep 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Feb 2021	Mar 2021	Apr 2021	May 2021	Jun 2021	Jul 2021	Aug 2021	Sep 2021	Oct 2021	Nov 2021	Dec 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	May 2022	Jun 2022	Jul 2022	Aug 2022
Define scope of curation coordinator role	■																							
Curation coordinator performs routine tasks		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Call for nominations for curation advisory board	Y												Y											
Curation board nominations due		■												■										
Curation board selection			■												■									
Announce curation board			Y												Y									
Prepare workflows for curation board			■												■									
Curation board in-person meeting				R	R											R	R							
Quarterly feedback from curation board								■			■				■		■			■			■	
Finalize technical roadmap	G																							
Technical implementation & testing		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
Preliminary report on curation requests													■											
Final report on curation requests																								■

Color Key

General Planning & Coordination	■ (blue)
Technical Work	G (green)
Outreach & Communications	Y (yellow)
Meetings	R (red)

DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

INSTRUCTIONS

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS.** Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

> ### SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS
> Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.
>
> ### SECTION III: SOFTWARE
> Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.
>
> ### SECTION IV: RESEARCH DATA
> Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). Explain and justify any prohibitive terms or conditions of use or access, and detail how you will notify potential users about relevant terms and conditions.

The Research Organization Registry (ROR) project provides all ROR IDs and metadata under a Creative Commons CC0 1.0 Universal Public Domain Dedication. We do not impose any restrictions or conditions on access to and use of registry data. ROR software is available under a MIT license and accessible via open Github repositories. Static content on the ROR website is licensed under a Creative Commons Attribution 4.0 International License.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

The ROR project will not assert any ownership rights over ROR IDs and metadata. We will not impose any restrictions or conditions on access to and use of registry data.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

ROR includes metadata for research organizations derived exclusively from public facts about these organizations, such as names and website URLs. No rights or permissions are required to include this information in the ROR registry. Some of the geographical data in ROR that indicates where organizations are located may have the potential to raise cultural or political sensitivities when countries/ territories are contested (e.g., Taiwan, Palestine). We plan to address such issues in two ways: first, by relying on trusted data authorities and standards for country data (e.g., ISO, GeoNames, United Nations) to represent this information as opposed to making our own judgment calls, and by developing appropriate curation workflows—as described in this funding proposal—so that we can make necessary adjustments to this data based on community input.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and the format(s) you will use.

A.2 List the equipment, software, and supplies that you will use to create the digital content, resources, or assets, or the name of the service provider that will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata or linked data. Specify which standards or data models you will use for the metadata structure (e.g., RDF, BIBFRAME, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content, delivery enabled by IIIF specifications).

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

SECTION III: SOFTWARE

General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

We will upgrade ROR's technical infrastructure to support independent curation of registry data. ROR needs to support basic curation functions so organization records can be added to, deprecated from, and updated in the registry. More advanced curation functions include adding or updating the metadata fields used in the ROR dataset, tracking changes to records over time and preserving record versions, and configuring an administrative layer for curators to review, coordinate, and complete curation tasks. The primary audience served by these upgrades include librarians, research managers, funders, university administrators, and other stakeholders who need to track research outputs by institutions and who will rely upon ROR data being comprehensive, up-to-date, authoritative, and responsibly managed.

A.2 List other existing software that wholly or partially performs the same or similar functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

ROR is one of several extant registries providing both identifiers and metadata for organizations. The most well known databases most similar to ROR are: GRID, ISNI, Ringgold, Crossref Funder Registry, LEI, PSI, and Wikidata. ROR is distinct from these because it is the only one with all five of these characteristics: (1) noncommercial; (2) open dataset; (3) open infrastructure; (4) community-led; (5) focused specifically on research organization affiliations. GRID, which is most similar to ROR and which provided the seed data for ROR's launch, is openly available under a CC0 license, but it is managed by a commercial entity, does not have an open API, and is not community-led. The other similar registries are either commercial, not openly available, and/or not focused on affiliations. ROR is needed to provide open community-led infrastructure for research organization affiliations.

Technical Information

B.1 List the programming languages, platforms, frameworks, software, or other applications you will use to create your software and explain why you chose them.

The ROR API and front-end search are the core components of the registry, storing and serving the entirety of the ROR dataset so it can be searched, matched, filtered, and viewed by both humans and machines. The ROR dataset is stored in Elasticsearch, a powerful, scalable, RESTful search and analytics engine. The ROR API retrieves data from Elasticsearch and serves it in JSON format. The ROR API is written in Django, a high-level web framework based on Python, which supports clean, loosely-coupled component architecture, allowing for rapid development. ROR's front-end search interface is built with Ember CLI. We have chosen this technology stack because it aligns with industry-standard state-of-the-art practices for open infrastructure projects.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

ROR's open API can be integrated into any system that collects institutional affiliation data, such as a journal submission system, grant application portal, or research data management system. Instead of collecting a free-text field, systems can make an API call to ROR's institutions list, e.g., a researcher based at UC Berkeley can type "UC Berk—" and be presented with dropdown options from ROR that match the entered text and then choose "University of California, Berkeley" from the list. This type of affiliation lookup allows for controlled and standardized affiliations as opposed to inconsistent free-text entries (e.g., "UC Berkeley," "UCB," University of California-Berkeley," etc.) ROR IDs are also interoperable with and are crosswalked to other identifiers: GRID, ISNI, Crossref Funder Registry, and Wikidata.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

Updating or adding records in ROR currently depends on these updates first being released in the GRID database, at which point ROR syncs the changes. This is a temporary workaround that was intentionally introduced in order for ROR to be able to launch its first registry prototype as early as possible using seed data from GRID. The IMLS project we are proposing is specifically aimed at setting up ROR to remove this dependency on GRID.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

We manage development processes and documentation on a public Github site: https://github.com/ror-community.

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous software your organization has created.

ROR is a new project and as such we have not previously created software on this project's behalf. ROR technical work is carried out by developers from two of ROR's collaborating organizations: Crossref https://www.crossref.org) and DataCite (https://www.datacite.org), which are leading global providers of infrastructure, metadata, and persistent identifiers for scholarly research outputs.

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

ROR software and source code are publicly available on Github under a MIT license at https://github.com/ror-community.

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

ror-api
ror-app

URL:

https://github.com/ror-community/ror-api
https://github.com/ror-community/ror-app

SECTION IV: RESEARCH DATA

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

A.1 Identify the type(s) of data you plan to collect or generate, and the purpose or intended use(s) to which you expect them to be put. Describe the method(s) you will use, the proposed scope and scale, and the approximate dates or intervals at which you will collect or generate data.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

A.7 Identify where you will deposit the data:

Name of repository:

URL:

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?