Organization: Research Organization Registry (ROR)
Primary mentor: Maria Gould, ROR Project Lead

. Project Title	Research Organization Registry: Curating a Global Registry of Affiliation Metadata
Description	ROR (Research Organization Registry) is an independent, noncommercial, openly licensed registry of persistent identifiers for research organizations. ROR IDs are designed specifically to solve the use case of capturing researcher affiliations in a standard and consistent way and connecting these affiliations to research outputs to enable more efficient discovery and tracking of research across institutions and funding bodies.
	Since its launch in 2019, the registry has grown to include open identifiers and metadata for more than 100,000 research organizations around the world. ROR IDs are recognized as the global research community's standard for identifying affiliations and are being widely adopted across core components of research infrastructure. ROR IDs are the recommended source of affiliation metadata in Crossref, DataCite, and ORCID, along with numerous publishing platforms and data repositories.
	ROR is unique among similar initiatives to provide identifiers for institutions because it is the only such registry that is completely open, exclusively focused on linking affiliations to research outputs, and developed for and by community stakeholders.
Problems/ Research Questions	• Review the metadata across the entire ROR corpus (including name variants/abbreviations, translations and multilingual formats, location data, related organizations, type categorization, and crosswalked identifiers) • Work with ROR Metadata Curation Lead to triage community-submitted issues, perform bulk data analysis and metadata QA, and develop, implement, and improve automation scripts for metadata analysis and

	production • Develop and improve automated and manual processes for the curation deployment process • Explore and analyze the relationship of ROR to sibling registries, such as the Crossref Funder Registry
Techniques	Consulting with ROR project team, data analysis, data mining of existing corpus, developing automated scripts, machine learning
Tools/ Languages used	Depends on the skills of candidate. Will work with json/XML exports and rely on scripting and/or manual checks. Programming language of choice (Python or R is preferred). Dashboard reporting tool like R Shiny, etc. or other forms for sharing output possible
Data	Description: ROR Registry records Data Type: JSON, tabular Data Size: metadata for 100,000+ research organizations
Outcome	Possible outcomes include: • Published analysis of existing or newly generated metadata • Presentation of data analysis to identifier communities and data science groups • Development of best practices and documentation for automated metadata workflows • An improved ROR corpus with a set of recommendations for improved curation practices
Milestone Timeline	To be worked out with mentor
References	ror.org Note: ROR is a cross-organizational and multi-stakeholder collaboration led by California Digital Library, Crossref, and DataCite, in conjunction with a broad global network of advisors. While this project's mentor is an employee of California Digital Library, the selected intern will work directly with joint project team led equally by Crossref, DataCite and CDL.