

NATIONAL SCIENCE FOUNDATION
Proposal Abstract

Proposal:1937136

PI Name:Ohno-Machado , Lucila

Proposal Title: Convergence Accelerator Phase I (RAISE): Knowledge Open Network Queries for Research (KONQUER)
Institution: University of California-San Diego
Abstract Date: 08/07/19

The NSF Convergence Accelerator supports team-based, multidisciplinary efforts that address challenges of national importance and show potential for deliverables in the near future.

The broader impact and potential societal benefit of this Convergence Accelerator Phase I project is developing new tools to allow researchers to conduct more accurate, informed research across multiple, disparate scientific domains. The initial efforts will provide research tools that address national concerns such as health; re-use and study of data that the public has already funded through federally awarded grants; and improving training in science, technology, engineering, mathematics, and medicine. The team includes partnerships with researchers from biomedical, social, geo-science, and climate science fields and integrates extensive expertise from data cyberinfrastructure efforts including: DataMed, a biomedical discovery index previously funded by the National Institutes of Health Big Data to Knowledge initiative; Data Discovery Studio, a geoscience discovery index funded by the National Science Foundation (NSF); and Pangeo, a climate science discovery and integration platform funded by NSF and NASA. During Phase I, the project will develop a prototype search engine, called KONQUER (Knowledge Open Network Queries for Research) that will connect these disparate data types and facilitate queries across the integrated datasets. The initial focus of the project is on biomedical, geological, and climate science fields; however, the technology can be extended to cover other scientific disciplines in the future.

Technological advancements have generated a large volume of data, but finding and analyzing those data in meaningful ways is often challenging. Real-world scientific questions cross multiple fields. For example, "Did the precipitation levels in California's Central Valley in 2016 cause an increase in the number of Valley fever cases?" To answer this question, requires data from health care, geolocation, and climate science. However, researchers are traditionally trained in only one field and are not familiar with the data resources of other disciplines, and even when they are aware of other data sources, the data may be structured very differently, all of which slows the process of discovery. KONQUER is envisioned to be a data discovery index capable of integrating various scientific fields and will use natural language processing tools (like a Google search) to decompose questions and retrieve information from the relevant data sources. To achieve this goal the project team will extend the DATS (DATA Tag Suite) metadata format so that geo/climate and science/biomedical datasets can be indexed in compatible formats, develop a pipeline for automated indexing, and develop a search engine that can query and rank the indexed data. The resulting KONQUER tool will accelerate and transform research and information retrieval so that new hypotheses and discoveries are possible across disciplinary boundaries.

NATIONAL SCIENCE FOUNDATION
Proposal Abstract

Proposal:1937136

PI Name:Ohno-Machado , Lucila

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.