

NATIONAL SCIENCE FOUNDATION
Proposal Abstract

Proposal:1936731

PI Name:Bodik , Rastislav

Proposal Title: Convergence Accelerator Phase I (RAISE): Linking the Open Knowledge Network to the Web with End-User Programming
Institution: University of Washington
Abstract Date: 08/12/19

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 to create tools that will make it easier for anyone, including non-programmers, to contribute to an open knowledge network, which is a nonproprietary, shared knowledge infrastructure that facilitates searches and queries of all publicly available data. An overarching thrust of NSF's 2019 Convergence Accelerator program is to create an open knowledge network (OKN) and this specific project's efforts have the potential to improve the development of knowledge networks in any topical domain. The project effort will build upon the team's existing research to enhance a programming tool that will allow people without programming expertise to acquire and use large, complex datasets from the web, so that anyone can contribute to an open knowledge network and help create a shared resource of broad value.

The motivation for the project is the need for researchers, decision-makers, and the public to have access to more thorough and current data, focusing first on data relevant to the broad fields of sociology, public policy, and economics. The project seeks to create tools so that social scientists can enrich an existing OKN containing government datasets with the dynamic view of the world presented by website data. This project will produce programming tools that allow social scientists (or any researcher) to create ontologies and link datasets without extensive programming training. To create a truly open OKN - that can be extended by untrained consumers, data providers, and other non-coders - requires highly learnable programming techniques. This project expands the team's prior advances in the field of learnable programming to build an OKN interface around cutting-edge technologies that merge programming languages, human-computer interaction, and programming-by-demonstration. Concretely, using the tool this project will develop a user will be able to encounter a target dataset on a website, the tool will demonstrate how to collect a small sample of the data, show the user how to annotate the data with its ontology (connections to other terms and ideas), then based on the user's input, the tool will write the program to extract the next trillion data points from the website and enter them into the OKN. The team also envisions that the tool can then suggest other datasets that may be appropriate to link, and the user can enter into a conversation with the tool to refine one or more linked scripts. The research team will build on their existing partnerships to ensure usability best practices for a range of target audiences.

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.