

NATIONAL SCIENCE FOUNDATION
Proposal Abstract

Proposal:1937160

PI Name:Baranzini , Sergio

Proposal Title: Convergence Accelerator Phase I (RAISE): A multi-scale open knowledge network for precision medicine

Institution: University of California-San Francisco

Abstract Date: 08/08/19

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 goal of this projects is to connect intersecting information across distinct domains and integrate it in a way that leads to innovation and new knowledge. For instance, if the US Treasury had access to private financial information from all domestic and foreign banks, financial markets, and even social networks, they might have been able to foresee the Great Recession of 2008. This project uses the same principle but is focused on computational tools and biomedical information with the potential to advance discovery across a range of biomedical disciplines.

Specifically, this Convergence Accelerator Phase I project will create a biomedical knowledge engine by integrating information from multiple biomedical databases. With the use of artificial intelligence, it will empower researchers with the tools to accelerate basic biomedical research as well as drug discover, and doctors with potentially unprecedented patient insights. The broader impact and potential societal benefit of this program lies on its ability to facilitate and democratize access to highly-specialized, yet publicly available information of biomedical importance. Currently, the majority of biomedical data sources are secluded within dedicated portals with minimal, if any, integration. This convergent project integrates wide-ranging expertise, from genomics, pharmacology and patient care, to computer science, deep data science, and epistemology. This team comprises research groups from academic institutions (University of California San Francisco), government (Lawrence Livermore National Labs), non-profit (Institute for Systems Biology) and commercial entities (Google), and during the project will engage with other organizations. While these groups have been collaborating (mostly in a pair-wise fashion) in projects with a smaller scope, this project intends to crystallize the team around a larger, common objective. The deliverables specified in this Phase I effort will contribute to the creation of a knowledge network (graph) composed of billions of concepts connected by biologically meaningful relationships. The graph will be open to the general public, accessible via manual or automated searches, whose content can be updated or modified, much like the world wide web is today.

Integrating vast amounts of information from multiple domains to an extensive knowledge network opens for the first time the possibility to computationally navigate the graph across disciplines that normally do not interact (like internal medicine and molecular biology). Furthermore, because the knowledge graph that will be developed contains domain-specific knowledge but not individual patient data, it does not pose privacy concerns. The intellectual merit of this project lies in the significant integration of domain expertise across team members and the leveraging of data resources that have

NATIONAL SCIENCE FOUNDATION
Proposal Abstract

Proposal:1937160

PI Name:Baranzini , Sergio

been created by other publicly and privately funded efforts. The project lays the groundwork for a resource that will potentially afford researchers and practitioners immediate access to the totality of all relevant "biomedical facts" - in the style of Google search, with the potential to change biomedical research, drug development and even the way medicine is practiced across the nation.

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.