

## Project Summary

In this research, we propose to design, implement, and evaluate Cedar: a CyberInfrastructure for Empirical Data Analysis and Reuse, to satisfy the requirements for Testbed I. Cedar is intended to be an open source information infrastructure architecture coupled with a data management policy mechanism that supports scalable and collaborative, qualitative and quantitative organizational research data collection, analysis, dissemination, and archiving. Our project involves the following components:

(1) *Infrastructure technology research and development.* Through the Hackystat Project, Principle Investigator (PI) Johnson has developed expertise in the development of open source collaborative systems for collection and analysis of quantitative data for software engineering research and experimentation. The Hackystat system and experiences provide a base for extension into qualitative data collection and analysis, as well as to a peer-to-peer network of federated servers.

(2) *Research on and development of policies and procedures for data privacy and dissemination.* PI Basili is leading a task force of software researchers with experience in developing and maintaining software engineering empirical data repositories with the goal of articulating prior problems and proposing improvements for management of future repositories. We will leverage this initial research and incorporate related research in privacy policies and technologies for integration into the Cedar infrastructure.

(3) *Research on and development of models and mechanisms for representation and integration of qualitative and quantitative information.* PI Pentland and PI Feldman have carried out a variety of research on the theoretical underpinnings of qualitative and quantitative empirical data and its appropriate interpretation. Cedar will leverage these insights with technological infrastructure for collection, analysis, and dissemination of empirical data according to narrative and network theories for representation and analysis of qualitative and quantitative data.

(4) *Case study evaluation of Cedar.* The four PIs (Johnson, Basili, Pentland, Feldman) have substantial prior experience in the design and implementation of case studies across a variety of application domains and organizational types. To test the validity of Cedar, and to understand its strengths and limitations, we will perform a case study with selected organizations involved in the DARPA HPCS program. As part of its evaluation, we will also develop Cedar-based curriculum materials to support education and technology transfer.

The intellectual merit of this research includes the application of novel data gathering and analysis techniques for the collection, integration, analysis, and dissemination of qualitative and quantitative data, and the application of this framework to a real-world organization and resulting evaluation.

The broader impact of this research includes the development of a sophisticated, freely available, open source software system for use by scientists for collection, analysis, integration, and dissemination of qualitative and quantitative data, along with associated curriculum materials to support education and technology transfer. As the University of Hawaii is a university with 75% minority students in an EPSCOR state, this project will provide novel research opportunities to underrepresented groups.