Project Title: **Operational data provenance and cybersecurity for anticipatory disaster communication built on mesh networks**

Project lead and contact details: Christina Bandaragoda, University of Washington; cband@uw.edu
Project partners and contact details: Graciela Ramirez-Toro, Interamerican University of Puerto Rico; Patricia Ordóñez, University of Puerto Rico Rio Piedras; Tim Sauder, Olin College
Partners on related projects: Fernando Rosario-Ortiz, University of Colorado Boulder; Amy Pruden, Virginia Tech

Project Summary: Technological advances have enabled massive quantities of data collection for studying earth system changes and hydrologic natural disasters (i.e., hurricane, flood, drought). These data sets still need to be leveraged at the community-scale to improve existing disaster response with anticipatory planning, which demands fundamentally new ways of storing, accessing and processing hydrologic data and community interaction with the information for resource management. Operational cyberinfrastructure that maintains data provenance and cybersecurity of real-time input-output, is needed in disaster-driven extreme event hydrologic science and related population health and environmental research. This requires a mode of knowledge transfer that extends beyond the scope of most domain-specific science programs and advances data ownership, security, privacy, and accessibility of shared public resources for real-time crowdsourced collection and distribution of information. Our project aims to provide project continuity between multiple hack events and research projects in order to bridge water, health, resource management (related to community resilience to natural hazards. We aim to develop and advance cyberinfrastructure disciplinary science by archiving and distributing real-time community crowdsourced information synthesized with baseline and current information, in usable formats needed by vulnerable communities to build their own capacity to prevent disaster.

This synthesis project will leverage ESIP lab support with National Science Foundation Rapid Response Research in Puerto Rico after Hurricane Maria (Project 1), National Institute for Health Big Data to Knowledge project building diversity and big data innovations (IDI-BD2K) in Puerto Rico (Project 2), open-source cyberinfrastructure for water research (Tool 1) and open source crowdsourcing (Tool 2) to prototype a communications system which supplies and distributes information on water, energy, food and medical resources. This proposed innovation builds community-scale capacity to anticipate needs, coordinate preparations for future events, and provide information in anticipation of recurring natural disasters (such as hurricanes, floods and droughts). The data archive developed in Project 1 will be used in rapid prototype testing in two successive data science hackathons: 1) the Mesh Hackathon (Event 1, coordinated by project 2) will initiate a prototype communications system with data privacy, security, and provenance principles using the Ushahidi) mobile crowdsourcing platform (Tool 2; then 2) Waterhackweek (Event 2, operated by Tool 1 developers) will improve usability of water, population health, and community needs information into the prototype. To inform the work at the hackathons, the ESIP lab project (proposed) will support human-centered design and testing of the data-driven mesh prototype with community water system operators and owners of private water data, including travel between Events 1 and 2 by Project 1 and 2 students and early career researchers, and sharing of code and research outcomes with the ESIP community.

Project Outline: Collaborating Project 1: A NSF RAPID project generating research products from the US Hurricane 2017 season will provide the drinking water, environmental baseline and population health data for a Hurricane Maria disaster use case. The NSF RAPID *Building infrastructure to prevent disasters like Hurricane Maria* (December 2017-November 2018) project is focused on designing, prototyping, and testing the research software and cyberinfrastructure needed to prevent drinking water related disaster related to loss of physical infrastructure that can follow hurricanes. The RAPID project (outlined in Figure 1) is providing access to 1) a public data archive of baseline Puerto Rico (blue row), including geolocated water sampling data post-Hurricane Maria, and population health surveys, with capacities for private protected information (red row). With design to provide easy

access, 2) scientific researchers could use these data products to investigate health impacts of disaster events related to contaminated drinking water. The software infrastructure depends on 3) advancing the integration of flexible data formats like the Landlab rastermodelgrid and Observation Data Model (ODM2), which are designed with metadata that captures space-time variations, and open-sourced software installed on CUAHSI JupyterHub Server. These components are designed to maintain data provenance of consenting data providers (e.g. household and utility water samples), research scientists, code developers, and project contributors. Best practices for maintaining confidentiality and cybersecurity will be disseminated to new users, used for educational purposes, with the aim to eventually integrate the data and tools for use in HIPAA-compliant computing environments for use in research. This RAPID partners with Interamerican University of Puerto Rico, UC Boulder and Virginia Tech to collect drinking water samples in Puerto Rico after Hurricane Maria.

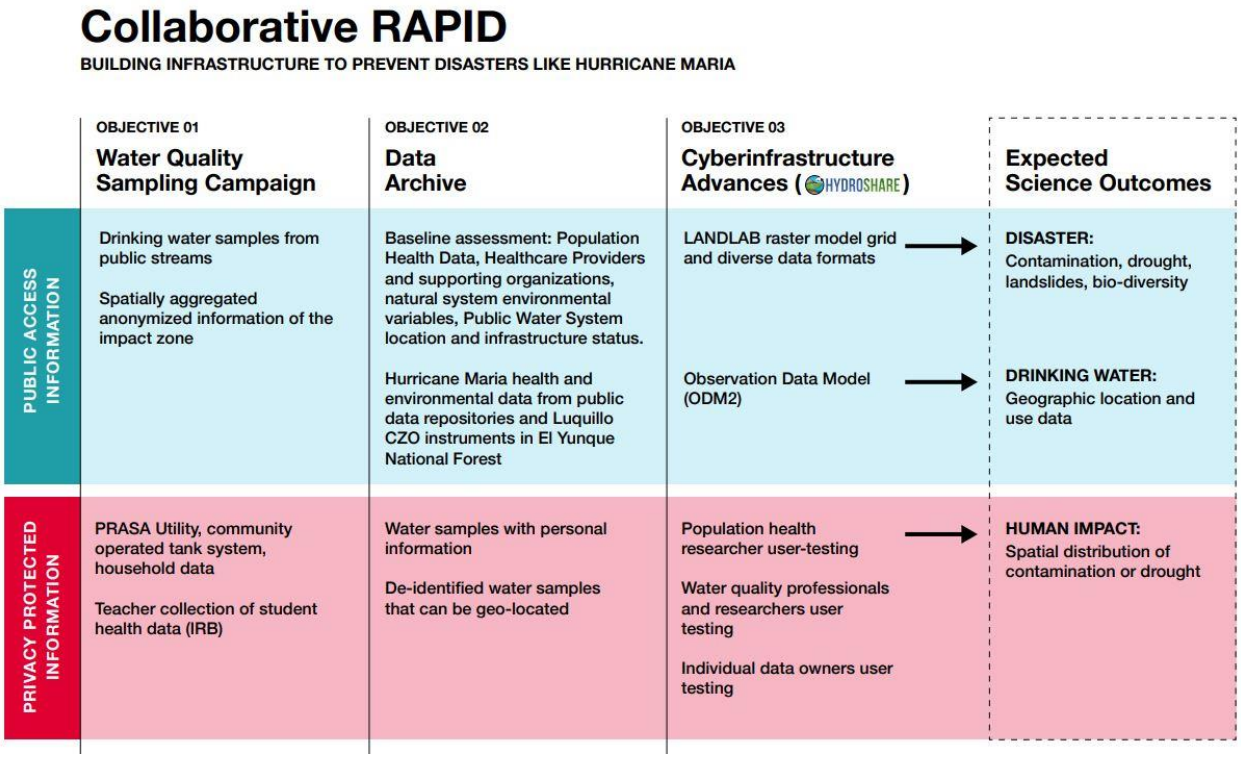

Figure 1. NSF Rapid Response Research (RAPID) project description for prototyping cyberinfrastructure developments for population health research.

Collaborating Project 2: The NIH funded *Increasing Diversity in Interdisciplinary Big Data to Knowledge (IDI-BD2K) in Puerto Rico* project is supporting the Mesh Hackathon (Hack Event 1 in San Juan, Puerto Rico, U.S. Territory) and related educational activities that enhance the training of a workforce to meet the nation's biomedical, behavioral and clinical research needs in Big Data Science. The IDI-BD2K Program at the University of Puerto Rico Río Piedras (UPR-RP) is focused on increasing the number of underrepresented researchers, and offer opportunities for faculty and graduate students in Puerto Rico to train Big Data research, expand their expertise, and thereby spur innovation in biomedical research on the island. The aim of the Mesh Hackathon is to develop a mesh communication network for coordinating emergency medical aid.

Tool 1: HydroShare is the online community repository that we will use to store, compute and share analyses for baseline and historic system data needed to develop scenarios and conduct contingency planning. HydroShare developers are supporting Waterhackweek (Hack Event 2 in Seattle, Washington, U.S.A). HydroShare is a cyberinfrastructure platform built to advance hydrologic science

by enabling the water-resources community to more easily and freely share products resulting from their research and/or data collection. HydroShare is freely accessible to everyone, but primarily was designed for water-resources professionals, researchers, educators, and others interested in water-resources data and tools. HydroShare is a system operated by The Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI), but has never been tested for integration with mesh network integration. HydroShare also includes tools (web applications) that provide users with a gateway to high performance cloud computing. With HydroShare you can: share data and models with colleagues; manage access to shared content; access, visualize, and manipulate a variety of uploaded hydrologic datasets; publish data and models and obtain a citable digital object identifier (DOI); aggregate resources into collections; discover and access data and models published by others; and use the web services application programming interface (API) to programmatically access resources.

Tool 2: Ushahidi is an open-source real-time crowdsourced data collection and data providing service which we will test for integration with the HydroShare platform using RAPID project data. Ushahidi offers products that enable local observers to submit reports using their mobile phones or the internet, while simultaneously creating a temporal and geospatial archive of events. Ushahidi is built as an API with a web client. It allows for custom survey creation, and the running of multiple surveys on a single deployment, embeddable maps and surveys, analytics, private deployments, and management of roles and permissions. It is built on the Laravel PHP web framework. It is open source under the AGPL license. We will explore archiving data collected in Ushahidi campaigns in HydroShare, and test using baseline data archived in HydroShare during a pilot Ushahidi campaign – thus advancing the cyberinfastructure prototype by networking the information between the two systems.

Project objectives, significance and impact:
Objective 1: Develop code to harvest RAPID Hurricane Maria related data from HydroShare for use within the Ushahidi mesh at the June 2017 Hackathon.
Objective 2: Develop mobile and print products in anticipation of the 2018 hurricane season in Puerto Rico. Test the usability of Ushahidi Mesh (by mobile phones) with communities in rural Puerto Rico with self-operated water systems in anticipatory planning for future hurricanes.
Objective 3: Develop code to maintain data provenance, privacy and security for contributors to the synthesized data archives. Test the usability for earth surface and health researchers at the November 2018 Waterhackweek.

Two major challenges will be addressed: 1) this project develops a qualitative process to build data science and data engineering skills with multi-disciplinary peer-mentoring in a project-based learning environment (with Project 1); 2) this project demonstrates emerging organizational methods for building sustainable research software with continuity, design, and innovation that bridges resources traditionally limited to project-by-project funding. The code, data and methods will be accessible to a broad and diverse water research audience using Github and multiple community repositories, while building data analysis literacy in specific academic disciplines, cultivating best practices of open data and sharing methods for both public and confidential data. The products of the proposed ESIP lab are expected to advance the standards of Accessibility in the Findable, Accessible, Interoperable, Reproducible (FAIR) standards. The data will be translated for use by communities participating in human centered design that provides formal control of privacy, confidentiality and sharing with secure online and print resources.

Description of key project steps and timeline:
June 29-30, 2018 Mesh network hackathon
August 2018 Usability and advertising campaign in remote southern region of Puerto Rico
November 5-10, 2018 UW Waterhack week
TBD, 2019 ESIP meeting presentation

Proposed start and end date: June 15, 2018 - May 31, 2019

Budget Requested: $9000
Budget Summary:

- Travel for two student/early career researchers currently conducting research from universities in Puerto Rico to travel to Seattle, Washington to participate in Waterhackweek: $4000
- Travel for one University of Washington PhD student (Seattle, WA to San Juan, Puerto Rico) to participate in BD2K Mesh Hackathon: $2000
- Human centered design and usability testing honorarium paid to community water system associations ($500 for three meetings; two hours per meeting; 20 participants per meeting): $1500
- Travel to ESIP meeting by project PI: $1500

Outreach: The target audience of this project are household and community data owners in Puerto Rico who anticipate future hurricanes. These community data owners would engage with academic and government scientists to generate and use data for resource planning in disaster situations (water, energy, food & human resources access and related infrastructure) critical for improving resilience to water related natural disasters. The hackathon student/early career researchers will conduct rapid prototype testing of the mesh communication network for usability with the community data owners. The community data owners will be asked to answer a survey addressing their technology acceptance and perception towards use. The student/early career researchers will share design diagrams and summaries to represent the information needs of the target users, the system requirements taken into consideration, mockups of the architecture and the proposed interaction diagram in Github. At the close of each hackathon, participants will update the github codebase, provide documentation for future use, and evaluate the product with regards to the project goals.

Project Partners (as applicable): Description of project partners (individuals and/or organizations) and their involvement: Dr. Christina Bandaragoda is coordinating the synthesis and cyberinfrastructure design, and will oversee the integration of water system data collected by 2017 NSF RAPIDs (including training for and collaboration with University of Colorado Boulder RAPID PI Fernando Rosario Ortiz and Virginia Tech NSF PIRE PI Amy Pruden). Dr. Graciela I. Ramírez-Toro, Director of the Center for Environmental Education, Conservation and Research of the Inter American University of Puerto Rico (CECIA-IAU) will oversee the collection and confidentiality protection of water system data collected by UC Boulder and Virginia Tech colleagues and made available to collaborating scientists, government, community and household data owners, supported by Olin College Professor of Design, Tim Ferguson Sauder, who is leading the human centered design and testing by Puerto Rico community users. Increasing Diversity in Interdisciplinary Big Data to Knowledge (IDI-BD2K) in Puerto Rico, co-PI Patricia Ordóñez, University of Puerto Rico, Río Piedras will direct the integration of NSF RAPID data with mesh networks at the Mesh Hackathon, with code development and data management support by Jim Phuong, University of Washington graduate student funded by the NSF RAPID.

Engaging members of the ESIP community: ESIP lab will be included as an event sponsor for the Mesh Hackathon and the Waterhackweek events. Project summary and links to code and research projects will be included in a ESIP newsletter article. Results will be presented at the ESIP 2019 Annual meeting. The project will be designed for scalability and usability to water-related disasters (flood and drought) by the global natural hazards and disaster preparedness communities. Advances in data provenance for data and model code reusability will be a general contribution for ESIP members coupling big data, earth surface models, and cyberinfrastructure platforms. ESIP members will be invited to apply to participate in and/or sponsor Waterhackweek projects.