

Proposal Clarifications

- **Please let us know if this work overlaps with any other funded work.**

This work does not directly overlap with any other funded work. However, elements of software and cyberinfrastructure that we will use are actively in development in other projects including HydroShare, ODM2, Landlab, Luquillo CZO, and NSF RAPID projects for post-hurricanes Harvey and Irma. We expect this project will utilize and inform the ongoing work of related projects, which we are leveraging with our strategically designed team. The Virginia Tech is conducting NSF funded investigations with study sites in Texas, Florida and Louisiana, focusing on groundwater wells, flooding hazards and pathogen analysis. This project expands that work by adding a study site with different characteristics, specifically surface water sources. Virginia Tech also has an NSF PIRE grant, which has good synergy with this project in terms of international focus on impacts of wastewater treatment plants to receiving environments and would financially support a graduate student to do field work and data analysis.

- **Identify the specific pieces of new or enhanced cyberinfrastructure that will result from this project.**

In addition to data collection and data hosting on HydroShare, this project has a new unique components, 1) application of ODM2 data model for samples of drinking water pathogens, 2) experimental design to incorporate the spatial distribution of public water infrastructure and impacts of loss of electricity using Landlab, and 2) design of research workflow that integrates ODM2, Landlab, and HydroShare, which have not been part of any other grant. Enhancing the accessibility of water quality databases through a public platform will provide an additional location for distribution of water quality data for health research. Links to the hydroshare repository will be added to uwaterstudy.org, enhancing the accessibility of research products.

- **How you plan to preserve this prototype cyberinfrastructure (such as the data collection kits) for future use?**

We will preserve the prototype through multiple community platforms used by the various domain scientists involved in the project.

Water Sampling Community: Methods for working with utilities and/or citizen science and the testing supplies needed (kits) will be described in plain language in the Virginia Tech supported webpage, uwaterstudy.org, with links to the project webpage at cuahsi.org, and detailed in publications. Online text will include water sampling campaign research details, issues related to data collection, including links to data, software tools, and the HydroShare data repository.

Natural Hazard and Engineering Community: The UW NEHRI RAPID facility will provide a website and on-line archive (<https://www.designsafe-ci.org/>) to be used to archive all information and research products generated from this research. This project can serve a case study of water sample collection, used by the RAPID facility to promote best-practices for post-event data collection. Online text will include research details, issues related to data collection, and links to data, software tools, and the HydroShare data repository. Including our project will extend the facility case studies beyond earthquake response, to include our examples of hurricanes, floods and water resources impacted by natural disasters.

HydroInformatics Community: HydroShare will be the central repository for the data collection. Users will be given instructions on websites listed above, to join the RAPID Group on HydroShare, whereby we can keep users updated on available research products and partner projects. [HydroShare platform code is available on Github](#). In addition to publication in HydroShare, Landlab code will be published in [Landlab Github repository](#), which is also distributed through [CSDMS](#).

Population and Public Health Community: Prof. Sean Mooney will lead preservation, development and distribution of our research products for the health research community. As the Chief Research Information Officer of UW Medicine, the Director of Biomedical Informatics at the Institute of Translational Health Sciences (ITHS) at the University of Washington, a PI of the National Center for Clinical Data To Health funded by the NIH, PI of an NSF funded project focusing on a bioinformatic community challenge focusing on data science in protein and gene function prediction (CAFA), he will direct the publication and use of the prototype information through these project and organizational websites. As part of these efforts, he is acutely interested and aware of the impact that environmental cyberinfrastructure, such as HydroShare, could have on clinical delivery systems and their IT infrastructure. We will develop communications for the community on integration of water, air and other environmental datasets into population health and clinical electronic health record systems for the purposes of care delivery, research and population health. We have proposals in review that could harness environmental datasets for predictive analytics and for clinical dataset integration. Finally, he will use his leadership within ITHS and UW Medicine to continue to support these datatypes in research surrounding human health.

Luquillo CZO and LTER Community: These research communities study areas of the El Yunque National Forest, which provides drinking water for approximately 20% of Puerto Rico's population. The [Luquillo Critical Zone Observatory \(CZO\)](#) supports a multi-disciplinary team of geoscientists studying critical zone processes. The [Luquillo Long Term Ecological Research Network \(LTER\)](#) conducts tropical forestry research in the Luquillo Experimental Forest. Our project collaborator, William McDowell, is a CZO PI and Climate rep for the LTER, and will work with co-PI Miguel Leon, CZO Data Manager, to preserve and make our research products available through these existing communication systems and websites. One goal of these collaborations with the USDA Forest Service and University of Puerto Rico is to provide a historical record of research for future scientists and students.

- **What is your plan is for dissemination of the data to other researchers in their community?**

We will begin to communicate about the data availability starting at AGU 2017 (New Orleans, Dec 11-15) where we will have a CUAHSI booth disseminated postcard advertising the RAPID project in Puerto Rico. We will also invite other researchers to Collaborate by joining the RAPID Group on HydroShare. Press releases developed at the University of Washington and Virginia Tech will be shared for distribution with partners on the project, including the [CSDMS](#) community and the UW Global Health Initiative. Publications resulting from this work will include information on joining the project information network, including a CUAHSI listserv dedicated to RAPID post-disaster water related research which will be included on the project webpage hosted by CUAHSI. Presentations will be delivered at various conferences, websites and peer-reviewed manuscripts. Target conferences include: Water Quality Technology Conference, [CSDMS](#) annual conference, and AGU 2018. Data and code for UW student projects and capstones will be made available through links and articles on the [Freshwater](#)

[website](#) and [UW eScience Institute website](#), which both include interdisciplinary water and data science communities.

- **What is your plan for preserving the privacy of potentially human-identifiable data?**

Virginia Tech holds IRB approval for several projects that are similar in nature in terms of human-identifiable data and presently has temporary approval for the proposed work. We will update our IRB approval specific to the Puerto Rico project upon award. Specifically, all resources with potentially human-identifiable data will be maintained as 'Private' on our team HydroShare resources. Only the project co-PIs and Senior Personnel will have access to ownership and editorial control of these private datasets after signing a Data Use Agreement with specific Terms of Use (provided by UW). De-identified versions of the data will be spatially aggregated to sub-watershed, census block areas, or regions, which will be reviewed by data stewardship and bioethics team members. After passing review for privacy, the de-identified datasets will be saved as public resources on HydroShare. We do not intend to specifically collect any identifiable information, but in the course of developing future partnerships, or in collecting geolocated data, we will follow the steps above to protect privacy of potentially identifiable data. Use of public utilities is also a potentially identifiable concern, so we will follow this procedure to preserve the privacy of potentially identifiable public drinking water treatment facility data.

- **Please indicate your willingness for someone in your team to attend the SI2 PI Meeting to share their experiences and your findings (likely through a poster session).**

We are willing to attend the SI2 PI Meeting. Co-PIs and senior personnel have existing grants from the SI2 which provides funding for traveling to DC to attend SI2 PIs meeting. We can utilize those funds to present results from several projects.

- **What metrics or other means will you use to measure the progress of as well as manage your project?**

Data collection will be a short-duration high-intensity effort with milestones that will be accomplished early on. Design of the system with health researchers and other users will be tracked through user surveys and user-design interviews. Use and traffic of the data through Google analytics on [uswaterstudy.org](#) and the CUAHSI project landing page. Several students and post-doctoral researchers will gain substantial experience stemming from this effort. For managing the project we will use Trello software, as well as CUAHSI processes for project management with scrum boards and sprint tickets. Active users will be tracked by using the Collaborate feature in HydroShare groups using methods we have developed for the Landlab group on HydroShare, which has resulted in tracking 200 Landlab users on HydroShare.

- **Is there a connection with citizen science and public participation in science research?**

This research proposal does not include funding for citizen science data collection, but we do intend to involve the public/citizen scientists in testing the usability of the software infrastructure. The time-sensitive data collection we prioritized in our proposal focuses on sampling and analysis for pathogen contamination from public water systems. These datasets are the most

difficult to acquire because they depend on participation of the government utility, involvement of a microbiologist in the sampling, and specialized lab equipment for analysis. As a pilot, we are targeting six water treatment utilities within a connected hydrologic system to investigate the water quality beyond the current safe drinking water regulations. Because each utility serves 1000-100,000 citizens, we feel this approach will have the highest scientific and societal contribution.

Standardized field collection and laboratory analytic methods have been established by the Virginia Tech team for the types of samples that will be analyzed. The scope of this project (current budget) is to collect 10 samples per public utility using their recommended sampling sites. For pathogen testing, Co-PI Amy Pruden's lab at Virginia Tech fulfills the need for capability and approval to do pathogen analysis. Utilities would have to partner with research institutions (e.g. EPA, Virginia Tech) with microbial equipment and microbiologists with the training to run these types of tests. The methods employed are intended mainly for research purposes (e.g., DNA sequencing and analysis) and not necessarily regulatory purposes according to the Safe Drinking Water Act. If we can not find six utilities that will work with us, we are planning to do a citizen science campaign where the citizens collect the samples, but work with microbiologists to run the analysis.

We are actively working to connect our work on this proposal to citizen science with public participation with the following activities (to date):

1. A University of Washington team including Engineers without Borders students using a UW Clean Energy Institute award to travel to the community of Jayuya, Puerto Rico to interview residents about the impacts to their health care based on lack of electricity. The team will explore how micro-hydro installations can be used to support rural residents how depend on medical devices to maintain their health. We are partnering with this team by adding on interview with questions related to their water source, filtration methods, and current water security. The team will inquire with local community leaders and PRASA water utility, if they would invite our RAPID team to sample their water. We anticipate this partnership will lead to our team initiating a field campaign in Jayuya, as one of our study locations.
2. We have submitted an application for funding for science communication to the UW Global Innovation fund, which we expect to hear about by Dec. 15, 2017. If funded, this award would pay for the design on science information provided with the science kits, including information on which water filtration systems are appropriate for various forms of contamination.
3. Colleagues from the Luquillo Long Term Ecological Research Network (LTER) are currently distributing water filtration in Puerto Rico, and have expressed interest in collaborating with our team to test household water for approximately 100 federally funded scientists and families living near the El Yunque National forest. We anticipate sharing our software infrastructure, findings, and methods with the network listserv, who has also expressed interest in seeking additional funding for water quality testing.