## Utah State University
# DigitalCommons@USU

Funded Research Records

1-22-2018

# WF-2331 NSF RAPID Building Cyber Infrastructure to Prevent Disasters Like Hurricane Maria

Jeffery S. Horsburgh
*Utah State University*, jeff.horsburgh@usu.edu

Follow this and additional works at: https://digitalcommons.usu.edu/funded_research_data

**Data Management Plan**

**Types of Data**

This project will compile, document and archive drinking water quality sample data with relevant associated environmental datasets, metadata, and contextual datasets from Hurricane Maria. Development of the required data storage and integration techniques, and processing methods/algorithms for enabling access to and reuse of the data by interested users is part of the intellectual contribution of this project. Development efforts on this project will be devoted to meeting these goals. The following are the types of data that will be assembled:

1. **Critical Zone Observatory (CZO) Data** will include environmental data related to Hurricane Maria, including meteorological data, precipitation, stream sensor and chemistry data, soil sensor data, and cloud monitoring data (ceilometer and cloud camera data) – as available given that some sensors were damaged or destroyed during the hurricane.
2. **Publicly Available Population Health Data** will include: 1) population health descriptive data from Institute of Health Metrics and Evaluation (IHME), 2) research agreements with Puerto Rican Healthcare stakeholders, and 3) health data needs for population health research stakeholders from Puerto Rican Hospitals and healthcare clinics.
3. **Drinking Water Sample Data** will include analytical results from water quality samples collected as part of this project from public drinking water sources. Analytical results will include laboratory analyses for water-borne pathogens and inorganic constituents.
4. **Geospatial Data** will include roads and road closure information, Safe Drinking Water Information System point locations of Community and Public Water Systems, locations of 69 hospitals in Puerto Rico, mudslide locations and landslide hazards obtained from the USGS, and storm deforestation rasters generated from satellite data. An effort will be taken to acquire information about electrical power availability in relation to hospitals and drinking water treatment and distribution systems, it is unknown at this time if these data may become available.
5. **Satellite Data** from NOAA satellite imagery, Landsat 8, WorldView-2 imagery, Digital Globe Satellite Imagery, and imagery from the National Weather Service. These data will also include night lights and storm path.

**Data and Metadata Standards**

We plan to organize and share the data we assemble within the HydroShare data repository (see plans for archival below). HydroShare makes full use of existing and emerging standards for sharing environmental datasets. All HydroShare resources (i.e, datasets, models, scripts, etc.) are described with metadata that conform to the Dublin Core metadata standard (DCMI, 2012), conform to a data model that is an implementation of the Open Archives Initiative's Object Reuse and Exchange (OAI-ORE) standard (Lagoze et al., 2008), and are stored on disk and packaged for download using the BagIt hierarchical file packaging specification (Boyko et al., 2012). These standards are well known within the library, information science, and digital archiving communities. HydroShare has also adopted standard file formats for the content files of known resource/data types. For example, HydroShare uses Version 2 of the Observations Data Model (ODM2) for time series data (Horsburgh et al., 2016), the Network Common Data Form (NetCDF) for multidimensional space/time datasets, ESRI shapefiles for vector geospatial data, and the GeoTIFF format for raster datasets. Part of the contribution of this project will be to extend HydroShare's capabilities of storing data derived from environmental samples using ODM2 for archival in HydroShare. This combination of standard data formats, standardized

metadata description, and standard packaging means that HydroShare resources are publishable and fully archivable.

## Policies for Data and Research Products

The goal of assembling Hurricane Maria-related data in HydroShare is to promote collaboration around, sharing, and reuse of these data. As such, all of the data we assemble as a direct result of this project will be publicly, and freely available in HydroShare using a Creative Commons License, where applicable. HydroShare is a data repository and collaboration environment. We anticipate that by curating and sharing Maria-related data in HydroShare we will enable users to not only discover and access the original data, but also to create derivative products through collaborative analyses and research that can also be shared within HydroShare.

When creating derivative products, HydroShare provides users with the choice to create public or private resources and public or private collaboration groups, accessible only to selected users, within which these activities can take place. Groups of researchers may wish to share data, model instances, or simulation results derived from the Hurricane Maria data within their group before they are published externally. To support this, authentication and access control have been fully integrated within HydroShare, and users can already choose how to share resources with other users or the larger community. Our experience has been that collaborative activities may result in multiple intermediate research products, only some of which may be considered publishable by the researchers. As such, HydroShare has functionality for creating formal, tracked versions of resources. Users can choose the Creative Commons License under which their resources are shared, and HydroShare has already established the facilities required to formally publish data, models, and simulation results, enabling individual researchers to select and publish their results as they see fit. Formally published resources are made immutable and receive a citable digital object identifier (DOI). Access to private resources and private research groups is at the discretion of resource and group owners. Final research results can be made freely and publicly available when they are deemed publication ready by the author. These collaboration capabilities make HydroShare an ideal location for publishing Maria data in a way that they can be cited, reused, and formally linked to derivative products.

All HydroShare resources have a landing page the displays the resource's metadata and contents, including attribution information (i.e., authors and contributors, funding agency credits, etc.) and a formal citation. HydroShare users must agree to a formal publication agreement prior to formally publishing a resource. This agreement was developed in collaboration with the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) and specifies the terms and conditions under which users can publish resources in HydroShare.

## Plans for Archiving Data

HydroShare will serve as the archival system for results from this project. Curated research products (e.g., datasets, models, etc.) published in HydroShare are citable for use in peer-reviewed journal articles, conference presentations and proceedings, and other formal publications using a formal digital object identifier (DOI). HydroShare and all of its attendant systems are hosted on fault-tolerant, enterprise-class servers dedicated to the HydroShare system and housed in the Renaissance Computing Institute's (RENCI's) managed, climate controlled, UPS-backed information technology facility, ensuring the reliability of the HydroShare system. All source code developed as part of this project will be openly shared in GitHub repositories associated with the HydroShare project.