9-30-2016

NRT: Graduate Climate Adaptation Research that Enhances Education and Responsiveness of science at the management-policy interface (Grad-CAREER)

Nancy Huntly
Utah State University, nancy.huntly@usu.edu

Data Management Plan

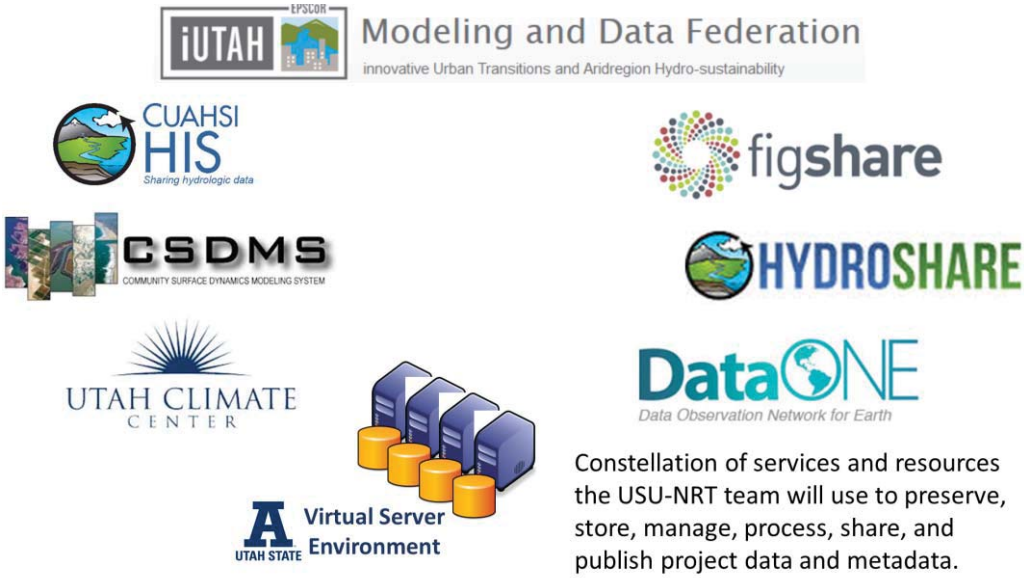
Data management is a core project component and represents a profound transformation of our graduate training program. All project faculty recognize the importance of empowering students engaged in data-intensive research with the skills to use emerging open science data and model cyberinfrastructure such as DataOne, Figshare, CSDMS, CUAHSI Hydrologic Information Services, HydroShare, Utah Climate Center, and the iUTAH Modeling and Data Federation. With many NSF programs and projects focused on cyberinfrastructure to support big data (e.g., EarthCube, iUTAH, HydroShare), this list will grow and our students are best served working primarily in this ecosystem. Thus, our data management plan purposely does not focus on building new cyberinfrastructure, but rather involves training project trainees and participants to work with their data in existing data management systems. Use will follow four core principles: (1) preserve data from corruption and loss; (2) transform raw observational data into fully tagged and annotated datasets that can be readily integrated for analysis; (3) follow established norms and requirements regarding human subjects data including de-identification and aggregation to protect confidentiality; and (4) ensure that all geographic, tabular, and image data, along with derivative works and models, are curated and shared with the scientific community for further use. Most cyberinfrastructure for open data curation includes a catalog that supports web services (e.g., OGC Catalog Service for the Web). Our project website will include a catalog page that lists and links to the data sets created by the project.

The research proposed requires the collection of existing data over a broad range of disciplines, including data on weather and climate (temperature, winds, precipitation), streamflow and water quality (discharge, water temperature, dissolved oxygen, dissolved organic carbon, nutrients, turbidity, total suspended solids), vegetation and other biota, land use and land cover, topographic data (1 to 10 m horizontal grid resolution), satellite imagery and air photos, surface and subsurface geologic maps, soil maps, land and water use data, maps of human infrastructure (dams, levees, roads, bridges, buildings, etc.), socioeconomics and demographics, perceptions from surveys and interviews, resource management and policy documents, GPS data, mathematical model codes, model input data, and course curricula. Some cyberinfrastructure may require adaptation, creation of additional features, or implementing and hosting at USU (e.g., DataOne member node). Therefore, we will additionally establish a virtual server environment at USU to provide: (1) a common file space for large data being manipulated within our campus high-speed network; and (2) a "sandbox" for students to work on virtual machines to experiment with implementing data cyberinfrastructure and developing enhancements. A high-level plan for the generation, validation, and delivery of data products follows:

Data Acquisition: All data will be compiled and maintained on Community Surface Dynamics Modeling System (CSDMS), CUAHSI, DataOne, or HydroShare facilities. Students will post their data (from computational analyses, instrument-based measurements, field notes, etc.) to a facility within 30 days of collection, with complete metadata.

Metadata: The open science data systems have developed rigorous metadata standards. We will use the "Qualified" Dublin Core metadata standard as our default and will extend these metadata requirements as needed for each of the open science data systems that serve as the appropriate outlet for each data type. Each file will contain unique metadata specific to that individual file (i.e., file- or resource-level metadata) and system-level metadata (relationships to other files), which allow it to be mapped for efficient search, discovery, and visualization.

Access and Archival: We will distribute data publicly, without cost, within 1 year of collection or once a corresponding paper has been accepted for publication. Data will be directly distributed via a project website and the online data systems listed above. Models and model components will be distributed via the NSF-sponsored CSDMS of HydroShare. Beginning Year 2, all downloaded, mirrored, and newly collected data will additionally be archived in the Utah Climate Center (UCC) data server in a GIS environment. The UCC has operated the server with dedicated IT staff for the past eight years (http://climate.usu.edu), shown below, and will continue to do so through the duration of this project under the supervision of UCC assistant director and co-PI Wang.



Constellation of services and resources the USU-NRT team will use to preserve, store, manage, process, share, and publish project data and metadata.

Educational Materials and Assessments: The project web site will also host and make publicly available the educational materials produced, including but not limited to (1) student tech reports, (2) courses materials and (3) video recordings of faculty talks and students presentations. In addition, the website will provide access to assessment tools used to evaluate student and faculty performance, as well as anonymized assessment results. Assessments will include existing performance-based rubrics for target competencies and professional practices. Data will include descriptive statistics for the NRT participants, as well as comparison statistics from the same instruments used by comparison groups independent of the NRT project. Posted reports will include both statistical and qualitative analyses of the comparative data, anonymized to protect the identities of both student and faculty participants.