3-14-2019

Collaborative Research: Predicting post-wildfire sedimentation of reservoirs: probabilistic modeling of debris flow generation and downstream sediment routing

Brendan Murphy
Utah State University, bpmurphy@aggiemail.usu.edu

Data Management Plan

Our project will generate and compile a vast amount data, spanning a wide range of data types and formats. Proper and efficient data management is essential to achieve our research and broader impact goals. Equally as important is the opportunity to empower our students and post-docs 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 members of our research team are best served working primarily within those platforms. Thus, our data management plan purposely does not focus on building new cyberinfrastructure, but rather involves utilizing existing data management systems and training students and postdocs to thrive within those systems.

We will follow three 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; and (3) 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).

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, sediment transport data, vegetation, land use and land cover, topographic data, satellite imagery and air photos, surface and subsurface geologic maps, soil maps, maps of human infrastructure, resource management and policy documents, GPS data, and mathematical model codes. Some cyberinfrastructure may require adaptation, creation of additional features, or implementing and hosting at USU (a DataOne member node and unlimited subscriber to an encrypted box.com server that can be made publicly accessible). A high-level plan for the generation, validation, and delivery of data products follows:

Data Acquisition: We will establish a common file structure on a local server. All project data (from computational analyses, instrument-based measurements, field notes, etc.) will be posted to the local server within 30 days of collection, complete with metadata. Within 6 months, all data will be compiled and posted to HydroShare, CSDMS, CUAHSI, GrADS, or DataOne facilities.

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 Community Surface Dynamics Modeling System (CSDMS) or HydroShare.

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 technical reports, (2) courses materials and (3) video recordings of faculty and student presentations.