

CDI Statement of Interest

Section 1. Project Administrative Information**Science Support Framework Elements:** Data, Data Management, Derived Information Products**Project Title:** National Stream Characterization: A Standardized Process and Database**Lead USGS cost center:** USGS Core Science Analytics and Synthesis**USGS Principal Investigators (PIs):** Daniel Wieferich (USGS-CSS) Denver Federal Center Bldg. 810, Box 25046, Lakewood, CO 80225; Phone: 303-202-4594; email: dwieferich@usgs.gov**Co-PIs:** Jeff Falgout (USGS-CSS; jfalgout@usgs.gov)Brad Williams (USGS-CSS contractor; bradwilliams@usgs.gov)

Scott Leibowitz (EPA-Safe and Sustainable Water Resources National Program;

Leibowitz.Scott@epa.gov)Marc Weber- (EPA-Safe and Sustainable Water Resources National Program; Weber.Marc@epa.gov)Dana Infante – (Michigan State University, NFHP; infanted@msu.edu)**Collaborators:** USGS National Water Quality Program, USGS NHD Program, National Fish Habitat Partnership, USGS Fisheries Program, USGS Climate Mission Area

Project Description and Deliverables: As research and management of natural resources shifts from local to regional and national scales the need for information of aquatic systems to be summarized to multiple scales is becoming more apparent. Experiences of four national stream assessment efforts across the U.S. Geological Survey (USGS), Environmental Protection Agency (EPA) and National Fish Habitat Partnership (NFHP) show that a growing percent of scientific research funds are being dedicated towards the collection and processing of national datasets vs. the analyses themselves. Building off these separate efforts, we propose to develop a standardized workflow to efficiently summarize national datasets of landscape information into two hydrologically and ecologically significant scales of local and network catchments of the National Hydrography Dataset Plus Version2 (NHDPlusV2). Using this method we also propose to summarize landscape information for variables identified as priorities by the national assessment groups and other USGS and EPA researchers. These variables will include climate, geologic, soil, land cover and hydrologic variables. These data will be made publicly available, and summarization code will be documented and released to the public following USGS and EPA standards. The standardized methods and data are planned to be used in future national assessments by the USGS and EPA to help insure consistency and comparisons of future assessment products. In addition, the workflow will include a process for summarizing and releasing newly available national datasets as made available.

Section 2. Estimated Budget

| Category | Requested Funds | Matching Funds |
|---------------------------------|-----------------|----------------|
| 1. Personnel | | |
| Scott Leibowitz (GS 14, 2pp) | \$0 | \$8,880 |
| Daniel Wieferich (GS 9, 2pp) | \$0 | \$4,840 |
| Jeff Falgout (GS13, 1pp) | \$0 | \$5,764 |
| Brad Williams (Contractor, 1pp) | \$0 | \$8,653 |
| Marc Weber (GS12, 1pp) | \$0 | \$2,364 |
| EPA Data Technician Support | \$33,960.30 | \$0 |

| | | |
|--|-------------|-----------------|
| Total Salaries: | \$33,960.30 | \$30,501 |
| 2.Travel | | |
| 6x of 1 personnel for face to face meeting | \$7500 | |
| 1x of 1 personnel travel for CDI | \$1500 | |
| Total Travel: | \$9000 | \$0 |
| | | |
| Total Direct Cost | \$42,960.30 | \$30,501 |
| Indirect Cost (17.17%) | \$7039.70 | |
| | | |
| Grand Total | \$50,000 | \$30,501 (~61%) |

Section 3. Project Summary

Scope: Several research and management activities now require information from regional or national datasets to be summarized into ecologically relevant spatial units. Past experiences in the aquatic community show these sorts of summarizations to be a daunting task, consuming a great deal of time and resources. This makes it unrealistic for many projects to process needed information on their own. Several national stream assessment efforts have worked to address summarization issues for two spatial units known to be ecologically and hydrologically relevant (i.e. local and upstream network watershed catchments) and plan to make data available, yet each assessment effort currently uses different data workflows to get similar information. This opportunity will allow us to compare differences in workflows, pulling best practices and most efficient techniques from each into a standardized workflow which will produce a common USGS/EPA/NFHP dataset. This process and resulting data will help advance comparable national and regional level analyses across and beyond our agencies while reducing duplicated efforts and addressing existing data processing issues.

Technical Approach: We will evaluate existing USGS, NFHP, and EPA national assessment workflows and summarization techniques for computing performance, and quality of data output. We will use these evaluation results along with expertise from USGS High Performance Computing (HPC) to develop a standardized, efficient and robust process for stream characterization utilizing open source and parallelized computing environments.

Project Experience and Collaboration: The PI team is comprised of members from 3 national stream assessment efforts that have experience with stream characterization including EPA StreamCat, USGS Aquatic Gap Analysis Program, and NFHP (Wang et al. 2011; Tsang et al. 2014). The HPC group specializes in high performance computing. The USGS NHD program will be included to help expose products to NHD communities and users, and to ensure that this work is applicable to current and future NHD efforts (e.g. development of the NHD High Resolution Plus). In addition, USGS climate and ecosystem missions will be engaged to ensure processed data layers will support USGS science efforts.

Sustainability: The workflow will be developed allowing for easy integration between current any future versions of the NHDPlus stream network and will be stored on the USGS high performance cluster for efficient access and running of future datasets of interest. Standardized methods and data products will be incorporated into workflows of several large national efforts within the USGS and EPA helping to ensure sustainability into the future.

Timeline: We anticipate the following milestones to be completed by July 2016: Establish consistent data summarization process, incorporate and test on the HPC environment, and finalize the list of official landscape variables to process.

We anticipate the following milestones to be completed by the September 30, 2016: Process variables using HPC, document and release final code-set and data products.