6-11-2018

Collaborative Research: Rivers and the Carbon Cycle: A Mechanistic Basis for Dissolved Organic Carbon Removal

Michelle Baker
Utah State University, mbaker@biology.usu.edu

DATA MANAGEMENT PLAN

Data and Materials Produced

Data products will include time series of observations from sensors, including water height, temperature, dissolved oxygen, pCO_2, fDOM, and pH in 36 streams. Observations derived from field measurements will include stream width, sediment size distribution, stream discharge, ^{13}C of CO_2 and CH_4, spiraling metrics, stream metabolism. Observations derived from field or laboratory experiments will include organic matter decay rates. Observations derived from laboratory analyses will include excitation/emission matrices (EEMs), concentrations of DOC, POC, dissolved N and P, total N and P, ^{13}C of DOC and DIC. Model code and model outputs are also anticipated to result from the proposed work.

Standards, Formats, and Metadata

The project will make full use of existing and emerging standards for sharing data generated during the research. In most cases datasets will be in tabular form and will be stored as .CSV files. Metadata descriptions will conform at a minimum to the fifteen element Dublin Core, as supported by Utah EPSCoR's cyberinfrastructure team. For example, general metadata will include the purpose, language, and contributors. Spatial metadata include the spatial extent, coordinates, and site name. Temporal metadata include temporal extent. Variable names and units are specific metadata for each observation.

Roles and Responsibilities

Baker will have overall responsibility for implementing the data management plan. Prior to data collection, project researchers will meet to discuss standard operating procedures and procedures for data collection, organization, quality control and archival. The lead research on any given data collection effort will generate a data collection plan. Each plan will include 1) identification of the types of data to be collected, 2) description of methods used to create the data, 3) identification of data formats that will be used to store the data, 4) description of the data to be created, 5) description of who will have access to the data during collection, and 6) how the data will be made broadly available after completion and publication. Metadata for each data collection effort will be included. Routine data management (such as instrument calibrations) will be the responsibility of the individual PI's research group. Student researchers and other personnel supported by the project will receive training on data management and data sharing via their home institutions. Data management and sharing are part of Utah State University's required responsible conduct of research course, for example.

Dissemination Methods

Datasets, models, and model outputs that are defined by data collection plans will be made publically available through a data repository at the time the work is published. In most cases, we anticipate data publication and archival through CUAHSI's HydroShare platform. Dissolved organic matter fluorescence data (excitation and emission spectra of PARAFAC components) will be openly shared via OpenFluor, an online public repository of published fluorescence spectra. Code for statistical models will be open source and made available through GitHub.

Policies for Data Sharing and Public Access

Our data sharing and access policy is based on past collaboration with Utah and Wyoming EPSCoR. Data sharing among project team members will be done via cloud sharing (e.g., Box) with access to data available as soon as possible after collection. Data may be used exclusively by the dataset

collector for 12 months as specified in the data collection plan. Datasets, models, and model outputs will be made publically available at time of manuscript publication if not sooner as described above.

Archiving, Storage, and Preservation

All three PIs have redundant backups for data collected in this project: the original computer on which the data were created, external hard drives, and cloud storage (DropBox or Box). The PIs will ensure that all project personnel have access to all the data in case of personnel turnover (for example from illness). HydroShare also allows sharing and storage of data products and metadata within research groups before publication of the data product. To the extent possible we will use these HydroShare products as they become available. Once data products are in final form, they will be submitted for public sharing and archival via HydroShare, which will assign the data product a digital object identifier and will be publically discoverable. HydroShare has the advantage that its development has been supported by the National Science Foundation and CUAHSI. Dissolved organic matter fluorescence data (excitation and emission spectra of PARAFAC components) will be openly shared via OpenFluor, an online public repository of published fluorescence spectra. Code for statistical models will be open source and made available through GitHub.