Data Management Plan

The proposed project will yield field and laboratory data and modeling results describing the roles that microbiota (i.e., benthic fauna, tidal wetlands, and SAV) play in carbon and alkalinity dynamics in two contrasting tidal tributaries of the Chesapeake Bay: Potomac and York. The project will also generate evaluation data for the Broader Impacts Program.

Expected data

(a) Water column data. Surface and bottom (where stratification is present) water samples will be collected at 18 stations in the Potomac River Estuary and 12 stations in the York River Estuary during four seasonal surveys for two years (eight surveys total). The samples will be analyzed in lab settings for DIC, TA, $p\mathrm{CO_2}$, and $\mathrm{Ca^{2+}}$; one third of the stations will also be analyzed for organic alkalinity and pH. Simultaneously, the following ancillary parameters will be measured in the field: temperature, salinity, pressure, dissolved oxygen, chlorophyll-a, turbidity, and PAR.

(b) Benthic fauna distribution data. Benthic samples will be collected at 27 tidal stations during all cruises. Generated data will include organism taxonomic identification, biomass dry weight, shell dry weight, and ash-free dry weight. In addition, all empty or fragmented shell hash will be weighed to obtain total dry shell weight in bottom sediments.

(c) Benthic fauna calcification and respiration rates. At a subsample of stations in each tributary, additional benthic sampling will be conducted in order to collect 10–12 individuals of the dominant benthic fauna taxon for lab experiments that will generate data on calcification and respiration rates.

(d) Tidal wetlands carbon and alkalinity lateral transport. Lateral transport will be studied at two tidal marshes in the York River Estuary, generating a range of data types. LIDAR imagery will be used to create topographic maps for each marsh watershed at 1 m resolution. Sensors for water level and current velocity will be deployed at the beginning of the project to provide continuous data for two years. Discrete water samples will be collected hourly over a single tidal cycle in each of the four seasons at each marsh over two years and analyzed for DIC, TA, DOC, POC, and $\mathrm{Ca^{2+}}$. The measurements results will be combined with continuous monitoring data and a 3-D modeling approach to generate data on lateral transport.

(e) Submerged aquatic vegetation fluxes. SAV field incubations will take place seasonally over two years at two sites in the York and two sites in the Potomac. The specific sites will be selected annually based on publicly available SAV distribution surveys. Water samples will be analyzed for DIC, TA, $\mathrm{Ca^{2+}}$, and a range of ancillary parameters (dissolved oxygen, pH, temperature, salinity, and PAR). After the incubations are complete, SAV above- and below-ground dry biomass will be determined. The final generated data product will include DIC and TA rates quantified per unit area and per unit SAV dry weight.

(f) Microbiota carbon and alkalinity flux maps. Small-scale estimates of carbon and alkalinity fluxes from benthic fauna, tidal wetlands, and SAV will be scaled up to generate tributary-scale coverage.

(g) Riverine fluxes. The fluxes of DIC, TA, and $\mathrm{Ca^{2+}}$ from rivers to the estuaries will be estimated using the WRTDS model from available non-tidal monitoring data.

(h) Estuarine biogeochemical data. Historical tidal monitoring data from the CBP and prior research efforts will be aggregated into a data product containing the following variables: temperature, salinity, dissolved oxygen, pH, TA, DIC, and $p\mathrm{CO_2}$.

(i) Numerical model output. Diagnostic box modeling and 3-D process-based modeling will generate a range of data outputs for each tributary between 1985 and the present for the box modeling and 2015–2014 for the 3-D model.

(j) Broader Impacts Program data. Evaluation data from surveys will be collected using SurveyMonkey or similar online software. All data will be stored by PI Harris and her technician

and will only be accessible to her team, following typical protocols for human subjects data and is intended to secure privacy issues.

Data storage and sharing during the project

Over the course of the research project, the field data, laboratory analysis results and model output will be hosted at each responsible PI's institution. A dedicated Penn State Microsoft SharePoint account with unlimited storage capability will also be set up for the project. The SharePoint site will facilitate collaboration between project participants via easy file-sharing between the institutions and will also serve as an additional backup for all project-generated data. Outside data access requests will be facilitated by the lead investigator and preliminary data will be provided to interested parties outside of project investigators upon request.

Broader Impacts Program evaluation data will be curated to protect the privacy and confidentiality of survey and focus group participants following human research protection policies. All downloaded files or scanned data will be downloaded onto a password-protected computer accessible only to the Harris team, who have received human subject protection training. All evaluation data will be collected anonymously to protect privacy of the survey participants and any demographic data that could lead to identification will be excluded. Any findings will be shared only in an educational setting associated with evaluation and improvement of the broader impacts portion of the study. Some lessons learned may be shared in educational presentations at meetings focused on the geoscience educational pathway. There is no human subjects research associated with this project; evaluation data are only used here for program improvement and development.

Data archiving

We will adhere to the requirement that any data we collect or any data and modeling products that we generate will be made publicly available within two years of the final forms of these data and products.

Within 18 months of work completion, all filed and laboratory data (a–e), will be permanently archived at the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Rivest will archive the carbonate system data and the benthic fauna rate data; Fantle will archive the Ca^{2+} data; Woodland will archive the benthic fauna distribution data; Gurbisz will archive the SAV flux data; and Herrmann will archive the tidal wetland flux data. We will work with the BCO-DMO staff to manage our field and laboratory data and prepare the appropriate supporting documentation for each data submission.

Scaled-up flux maps, historical data products (f–h), and box model output will be permanently archived by Herrmann at the institutional repository at the Pennsylvania State University, ScholarSphere, with permanent DOIs. Output of the 3-D model simulations (i) will be permanently archived by St-Laurent at the Virginia Institute of Marine Science institutional repository, W&M ScholarWorks, with a permanent DOI. Institutional repositories are repository services produced by university communities, enabling university researchers to share their work on a worldwide scale and ensure long-term preservation and ongoing access to the data. The repositories provide assistance in complying with data management planning and data sharing requirements and enable researchers around the world to easily locate archived materials without subscriptions or passwords using unique DOIs. Links to the ScholarSphere and ScholarWorks DOIs will be made available on the project's BCO-DMO archive.

Broader Impacts Program evaluation data will be kept indefinitely by the Harris team on password protected computers with server backups following University of Maryland Institutional Review Board protocols. These protocols differ from those for natural science data and measurements to meet federal and regulatory requirements for human participants and align with best practices that respond to individual requests from other researchers to share data in a way that aggregates and de-identifies participants to insure privacy and confidentiality.