High-Resolution, Interagency Biosurveillance of Threatened Surface Waters in the United States

CDI FY19 Full Proposal

Lead PI Information

PI Name: Sara L Eldridge
PI Email Address: seldridge@usgs.gov
PI ORCID: 0000-0001-8838-8940
PI Organization: USGS Wyoming-Montana Water Science Center
PI Mission Area: Water Resources
PI City, State: Helena, MT

Plain Language Summary

Timely, up-to-date information concerning biological hazards that threaten the health of humans and aquatic habitats is critical for accurate detection and to minimize negative outcomes. If funded, we will create an automated and customizable tool to process and visualize new environmental data, as they become available, onto an existing online map of the Greater Yellowstone Ecosystem. This map will allow users to conduct risk assessments of pathogens and aquatic invasive species. Such an enhanced risk map will improve decision making and short-term planning by both natural resource managers and members of the public. Our data-processing tool could be modified to benefit users across USGS and within national parks and wildlife refuges and land managers across municipalities, states, tribes, and the private sector.

Financial Information

Total Requested Funds: 49979
In-Kind Matching Funds: 43800

Project Information

Project Description: To develop a pipeline for analyzing real-time and near-term eDNA data collected from MBARI robotic environmental sample processors at USGS streamgages and NOAA weather data to visualize and assess risk of aquatic invasive species, fish pathogens, and human pathogens based on Bayesian hierarchical occupancy models.
List of Anticipated Deliverables: script for data science pipeline, enhanced risk map, USGS Data Release
SSF Element 1: Analysis

SSF Element 2: Processing
SSF Element 3: Applications

Collaborators

	Name	**City**	**State**	**Organization**
Co-PI	Elliott Barnhart	Helena	MT	USGS WY-MT Water Science Center
Co-PI	Adam Sepulveda	Bozeman	MT	USGS Northern Rocky Mountain Science Center

High-Resolution, Interagency Biosurveillance of Threatened Surface Waters in the United States

Plain Language Summary

Timely, up-to-date information concerning biological hazards that threaten the health of humans and aquatic habitats is critical for accurate detection and to minimize negative outcomes. If funded, we will create an automated and customizable tool to process and visualize new environmental data, as they become available, onto an existing online map of the Greater Yellowstone Ecosystem. This map will allow users to conduct risk assessments of pathogens and aquatic invasive species. Such an enhanced risk map will improve decision making and short-term planning by both natural resource managers and members of the public. Our data-processing tool could be modified to benefit users across USGS and within national parks and wildlife refuges and land managers across municipalities, states, tribes, and the private sector.

High-Resolution, Interagency Biosurveillance of Threatened Surface Waters in the United States

Proposal Narrative

Scope

Goals and objectives

The goal of this proposed work is to develop a user-friendly script for processing large environmental datasets through sequential, chained steps (i.e., a pipeline), to conduct and visualize risk assessments of pathogens and aquatic invasive species (AIS) in threatened surface waters. We propose a Bayesian-based analysis of real-time and near-term NOAA weather data and environmental DNA (eDNA) data collected from robotic samplers at USGS streamgages to graphically visualize AIS and pathogen risk at the stream reach scale. The primary product will be a publicly-available digital map of the probability of AIS or pathogen presence in surface waters within the Greater Yellowstone Ecosystem, developed in cooperation with the Greater Yellowstone Coordinating Committee, that can be used to establish human health and ecological risk assessments. Additional products will include a documented workflow (pipeline) that can be applied across USGS, including Web Informatics and Mapping, the Nonindigenous Aquatic Species (NAS), and the StreamStats groups, and other agencies, and that will demonstrate the utility of the data science pipeline for general use in environmental biosurveillance. Key outcomes will be interagency data sharing for improved capacity for ecological forecasting and risk assessments of biological hazards and a data science pipeline to be shared with USGS and cooperators for integrating and visualizing similar data streams for various biosurveillance applications.

Background

Globally, aquatic ecosystems are threatened by multiple stressors, such as pathogens, invasive species, cyanotoxins, or harmful algal blooms. These stressors vary in distribution and intensity and are driven by human activities, climate change, and natural variabilities that affect ecosystems at different temporal and spatial scales. To support comprehensive, accurate assessments of complex ecosystem stressors, biological observations matching the scale and quality of *in-situ* physical and chemical measurements are essential. This need for high-resolution biological observations can be provided with applications of high-throughput environmental DNA (eDNA) detection techniques that generate vast pools of data relevant to interpreting changes in ecosystem health (Baird and Hajibabaei, 2012). Moreover, recent advances in high-performance computing, remote sensing, and geospatial tools have accelerated the adoption of integrated data streams requiring non-traditional analysis and visualization methods and the development of novel strategies for aggregating big data streams and disparate data types to study multiple stressors at the ecosystem scale and across different scientific disciplines. These diverse and often large data sources provide us with increased information at a resolution never previously realized (Dafforn and others, 2016; Van den Brink and others, 2016).

The opportunity now exists to develop a platform using genetic information for environmental monitoring at biologically-relevant scales, which is both time- and cost-efficient and that offers significant increases in the immediacy, accuracy, and quantity of information. In the early 2000s, the Monterey Bay Aquarium Research Institute (MBARI), a non-profit oceanographic research center, pioneered a portable robotic environmental sample processor (ESP), which collects and processes water samples *in-situ* and applies molecular probe technology remotely (Scholin and others, 2017). (https://www.mbari.org/technology/emerging-current-tools/instruments/environmental-sample-processor-esp/). The ESP has been used to detect toxin-producing harmful algae such as *Alexandrium*, a cause of

paralytic shellfish poisoning, and *Heterosigma*, which is deadly to fish. NOAA is currently expanding the ESP capability to enable near real-time detection of bacterial species that are known human pathogens and indicators of fecal contamination (https://www.nwfsc.noaa.gov/research/divisions-/efs/microbes/harmful_algae/esp.cfm).

Scientists from the USGS Wyoming-Montana Water Science Center (WYMT-WSC), USGS Northern Rocky Mountain Science Center (NOROCK), and the USGS Upper Midwest Environmental Science Center are collaborating with MBARI to incorporate ESPs at USGS stream gaging stations with the goal of providing near real-time DNA biosurveillance of U.S. rivers. The USGS is the Nation's leader in monitoring hydrologic conditions through its streamgage network (Olson and Norris, 2007). Integrating biosurveillance (AIS or pathogens) monitoring into the streamgage network in streams and lakes would provide real-time and near-term (5-30 days) collection of high-resolution physical, chemical, and eDNA datasets that can be used to create a near-term forecasting system for water resource quality and protection. As a test of using the MBARI ESPs at USGS streamgages, Elliott Barnhart (WYMT-WSC) and Adam Sepulveda (NOROCK) deployed second-generation ESPs at existing streamgages on the Yellowstone River adjacent to Yellowstone National Park in 2018. ESPs collected approximately 300 water samples every 3-12 hours for 42 days and preserved the filters. Samples were analyzed for the presence and quantity of *Tetracapsuloides bryosalmonae* (a myxozoan parasite of salmonid fish), invasive smallmouth bass, and *Escherichia coli* (a human pathogen). Results of these preliminary studies are building the foundation for establishing risk and early detection of disease, invasive species, and related water-quality issues, giving resource managers time to protect human and ecosystem health.

Occupancy models are the standard for analyzing presence-absence data to estimate the spatial distribution of a species (MacKenzie and others, 2018). These models are designed to predict the probability of detection, accounting for the possibility of detection errors, and they are useful for estimating the probability that an organism is present when it has not been observed or captured. Surveys of eDNA are designed to accommodate spatial or temporal heterogeneity in the occurrence of eDNA at a sampling location by the collection of multiple subsamples, because samples taken at the location occupied by a species may not necessarily contain eDNA of that target species. Therefore, presence-absence predictions based on eDNA data are more complicated by the fact that multiple subsamples of target organism DNA taken at a given location are tested, and each subsample may detect the organism, if present, with some probability. Hence, eDNA data structure requires an additional level in the hierarchical model to account for the probability of detecting the target organism through PCR. The class of multiscale occupancy models, developed by Nichols and others (2008), provides a useful framework for the analysis of data collected in eDNA surveys, in that these models account for the multiple nested levels of sampling (Dorazio and Erickson, 2017) that characterize eDNA surveys. Also, the Bayesian approach to inference is useful when analyzing data collected in eDNA surveys because the primary targets of inference may include latent states of species occurrence. Also, predictions of species occurrence at unsampled locations may be needed, particularly when decisions related to the management or conservation of a species are required. As an added benefit, Bayesian models provide a tool for updating data interpretations in the evidence of new data. So, Bayesian analyses are ideal for generating models from continuous data streams.

To fully realize the potential of integrated high-volume eDNA and physical data streams offered by adding MBARI ESPs to USGS streamgages, thereby continually generating new data to add to knowledge from existing datasets, statistical methods and computational software for real-time biosurveillance and near-term forecasting are needed to ensure timely implementation of appropriate management responses. An emerging challenge in collecting, integrating, and interpreting big data streams is the development of reproducible workflows and modeling pipelines. Structuring work flow by producing a data science pipeline is crucial for translating large datasets into meaningful interpretations efficiently and cost-effectively and for model refinement. The primary components of a data science pipeline include data acquisition; data preparation, cleaning (address missing or erroneous data), and quality-assurance evaluation; data exploration and visualization to identify significant patterns or trends; data modeling; interpretation; and model refinement. A flexible and customizable data science pipeline is

a simplified, reproducible, and effective means for adding new data to existing analyses following each sampling event for continual model refinement, improved accuracy of results, and greater predictive power, which gives resource managers the most up-to-date, high-quality information. Machine learning and evaluation methods can further enhance the predictive power of models through algorithms and statistical tools that build mathematical models to make predictions or decisions without being explicitly programmed to perform the task (Koza and others, 1996).

A critical component of the data science pipeline is translating data streams into functions with broad user access through data visualization tools. Data visualization in the form of interactive maps or a graphical user interface makes complex datasets more easily understandable, facilitates communication to more diverse data users, and bolsters informed decision making. Data visualization is currently used in risk assessments for AIS introduction at the stream reach scale in the Greater Yellowstone Ecosystem (GYE) with the USGS Greater Yellowstone Aquatic Invasive Species Explorer web interface (https://gagecarto.github.io/aisExplorer/). GYE managers currently use this web interface for prioritizing biosurveillance efforts, but more work is needed to upscale the interface to include relevant eDNA, river flow, and weather data streams.

Relevance to Departmental, Bureau, and CDI Priorities

This proposed study represents an interdisciplinary collaboration to leverage existing resources to improve access and functionality of existing data streams by creating a novel, but proven and customizable, method for analyzing and interpreting eDNA, streamflow, and NOAA weather datasets to inform resource managers and the public of the risk of AIS, pathogens, and other surface water hazards. Understanding and predicting invasive species is a key Departmental priority, and this proposed work is directly relevant to the USGS Director's FY18 Bureau Annual Guidance related to "Integrated Predictive Science Capacity," to "develop and deliver powerful new products and services that provide: 1) vulnerability detection and assessment, 2) prediction and forecasting, 3) early warning, and 4) decision support at the scale of decisions." The study report titled, "Future Water Priorities for the Nation: Directions for the U.S. Geological Survey Water Mission Area" (National Academies of Sciences Engineering and Medicine, 2018) highlighted the transcendent nature of eDNA and calls for improved analysis and visualization tools. The report also stresses that data analytics, such as the development of a data science pipeline to analyze eDNA and related data, also will be a critical area to "support improved scientific understanding, development of improved water quality models, and interdisciplinary model integration."

This proposed work also is closely aligned with the CDI Science Support Framework (SSF; https://my.usgs.gov/confluence/display/cdi/CDI+Science+Support+Framework) through the application of novel computational and analytical technologies. Our development of a data science pipeline and visualization of the data with a risk assessment map conceptually flows vertically through the SSF from a base of Monitoring and Assessment to Knowledge of Earth Systems, including aspects of Data; Processing, Analysis, and Sharing within the Scientific Data Life Cycle; to Web Services; Applications; and Information.

Further, this proposal aligns with the FY19 Topical Emphasis described in the current CDI Request for Proposals. Specifically, the CDI executive sponsors are encouraging proposals that address development of "tools and methods for biosurveillance of emerging invasive species and health threats." This will be the primary outcome of the study proposed here.

Outcomes and Deliverables

Development of a user-friendly, documented pipeline for processing MBARI ESP eDNA data, USGS streamgage data, and NOAA weather data through sequential, chained steps, and delivering risk assessments of pathogens and aquatic invasive species (AIS) in cooperation with key stakeholders is highly relevant to Departmental and Bureau priorities as described above. The methods, workflow, and

data integration and visualization framework are scalable and transferable, with acceptable modifications, to other aquatic plant, animal, or microbiological species for which eDNA detection methods have been developed and at local to national scales where mapping and forecasting are required. If funded, the work proposed here will focus on invasive species, a fish pathogen, and a human pathogen. But, the resulting data science pipeline may be customized for use in other applications, such as detection and monitoring of harmful algal and cyanobacterial blooms (HABs), zoonotic disease vectors (mosquitos, ticks), and their relationship with human health and safety, natural resource monitoring and assessment, and associated processes, such as climate change.

Anticipated deliverables include: 1) a pipeline for incorporating eDNA and NOAA weather data visually onto the Greater Yellowstone Ecosystem AIS Explorer map (https://gagecarto.github.io-/aisExplorer/) as these data become available to guide early detection of AIS and pathogen exposure risk, 2) a USGS Data Release and journal publication describing the pipeline steps for data acquisition, data preparation, data exploration, statistical analyses, and visualization, and 3) a formal presentation at the CDI workshop and presentation(s) during the monthly CDI conference calls with the goal of training and encouraging others seeking to implement the pipeline for similar biosurveillance projects. We also will collaborate with the USGS Web Informatics and Mapping, the Nonindigenous Aquatic Species (NAS), and the USGS StreamStats groups to enable them to integrate this pipeline in their online map resources.

Technical Approach

We propose to develop, implement, and document a pipeline, a user-friendly script of sequential, chained steps, to conduct and visualize risk assessments of pathogens and aquatic invasive species (AIS) in threatened surface waters using eDNA data collected from MBARI ESPs deployed at USGS streamgages, streamgage data (i.e., flow dynamics), and relevant weather data from nearby NOAA weather stations. Specific eDNA targets will include *T. bryosalmonae*, invasive smallmouth bass, and *E. coli* in samples collected from the Yellowstone River in 2018. Besides eDNA, the MBARI ESP also can collect samples for chemical analyses, so our proposed pipeline may be customized for use with different data types, such as algal and cyanotoxins or other chemical hazards. The overall workflow steps are as follows: 1. synthesize eDNA data, streamgage data, and weather station data; 2. employ USGS Bayesian hierarchical occupancy modeling to estimate target DNA detection probabilities and identify relations between eDNA and other data streams; 3. use forecasted streamgage and weather station data to predict eDNA occurrences and compare these results with observed values; 4. create a pipeline (defined process of sequential steps) to incorporate eDNA data, streamgage data, and weather station data into on-line visualization tools in cooperation with the Greater Yellowstone Coordinating Committee (state and federal land and natural resource managers) using observed and predicted eDNA occurrences as a function of streamgage and weather data; and 5. create and document a pipeline for automating these steps to process incoming data based on CDI Community of Practice standards.

Using the dataset from approximately 300 eDNA samples collected by MBARI's EPSs at two USGS streamgage sites on the Yellowstone River in 2018, we will develop spatiotemporal, multiscale occupancy models to identify how changes in streamflow dynamics and weather data are associated with the prevalence of our selected eDNA targets. These models will include parameters that specify the probability of target occurrence at study locations and conditional probabilities of occurrence in the collected samples. Model fitting procedures and assumptions will be similar to those described in Dorazio and Erickson (2017). Briefly, results of eDNA presence-absence determinations by qPCR will be used to compute detection matrices. We will then use a new, computationally expedient, R package (unnamed) (R Core Team, 2017), recently developed in collaboration with Montana State University (MSU), to fit multiscale occupancy models without, then with, the covariates listed in **table 1** using Bayesian computation methods. A basic multiscale occupancy model is already implemented in the R package described in Dorazio and Erickson (2017), but real-time surveillance and forecasting require faster computation and the ability to handle spatial and temporal components. Moreover, the Dorazio and Erickson (2017) package is no longer updated, because the lead investigator has retired. Our new R

package is based on a specific Markov chain Monte Carlo algorithm (Brooks and others, 2011), so that all calculations can be completed in R without need of additional software. Estimated relations between the probability of eDNA target occurrence in samples and the covariates can then be assessed. Given that our eDNA targets are aquatic organisms, we will fit linear regressions of water temperature as a function of air temperature from NOAA weather stations using the weekly average data as described in Al-Chokhachy and others (2013) for the Yellowstone River. Our new R package also enables model comparisons and predictions, so we will compute estimates of eDNA occurrence or detection probabilities for each target, location, and sample, and evaluate the models' predictions of eDNA occurrence as a function of covariate data streams. Comparisons between predicted and observed values during on-going field data collection, and across different environmental and hydrological conditions, will enable continual model refinement. We also will conduct sampling design power analyses, so that stakeholders can clearly assign confidence to any decisions based on past, current, and incoming eDNA data.

With our eDNA occurrence predictions, we will display near real-time and near-term risk assessments in a graphic-user interface format based on discussion and feedback from the Greater Yellowstone Coordinating Committee (GYCC). One option for data visualization will be to integrate our results with the USGS AIS explorer (https://gagecarto.github.io/aisExplorer/), which shows invasive species introduction risk in the Greater Yellowstone Ecosystem. Individual data streams and a composite risk assessment based on these data also can be presented to users with an R-based Shiny app, an interactive analysis and visualization web tool (Chang and others, 2018) that functions similarly to the NOAA HAB forecast maps and USGS StreamStats website (Ries and others, 2008). These tools will show users the observed versus expected occurrences of eDNA targets as a function of current streamflow and weather conditions.

The key component of this proposed work is to structure the workflow described above into a flexible, customizable, and fully reproducible data science pipeline for automatically incorporating new data into our occupancy models following each eDNA sampling event. We propose to model our pipeline after that used by the NOAA Great Lakes Environmental Research Laboratory, which combines *in situ* microcystin monitoring data (collected by MBARI ESPs), twice-weekly forecasts for *Microcystis* blooms as shown on the Lake Erie HAB Bulletin, and the Experimental HAB Tracker forecasts (https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/habTracker_about.html).

Table 1 Covariate data streams and associated data-collection intervals for use in proposed multi-scale occupancy models for predicting the occurrence of pathogen and AIS eDNA targets

MBARI ESP eDNA data		**USGS streamgage data**		**NOAA weather data**	
Variable	*Time interval*	*Variable*	*Time interval*	*Variable*	*Time interval*
DNA presence	3 hr., 12 hr.	Water temperature	1 hr.	Air temperature	1 hr. – daily
		Discharge	1 hr.	Precipitation	1 hr. – daily
		Air temperature	1 hr.	Snow depth	1 hr. – daily
				Wind	1 hr. – daily
				Solar radiation	1 hr. – daily
				Cloud cover	1 hr. – daily

Project Experience and Collaboration

The project team is composed of interdisciplinary USGS and collaborating scientists with experience and interest in community-level aquatic ecology, water quality, and developing new tools for the early detection of pathogens, cyanotoxins, and aquatic invasive species. Sara Eldridge, Microbiologist with the WY-MT Water Science Center, will serve as the lead PI for the project. Elliott Barnhart, Research Hydrologist with the WY-MT Water Science Center, and Adam Sepulveda, Research Zoologist with the USGS NOROCK, will serve as co-PIs on this project. Barnhart and Sepulveda have been collaborating with MBARI since 2017, and they have secured funding in successive years, primarily from

the USGS Innovation Center, to successfully demonstrate that the MBARI's ESP has the capacity to collect real-time data necessary to detect disease-causing organisms and invasive species in and around Yellowstone National Park, the San Justo Reservoir, California, and in Grand Teton National Park, Wyoming. This team also collaborates with researchers at Montana State University to develop methods and tools for biosurveillance and near-term forecasting systems for the quality of water resources. Previous work has been featured in Wired magazine several times (for example, https://www.wired.com-/story/the-hunt-for-the-brain-eating-amoebas-of-yellowstone/) and covered by local reporters. Multiple manuscripts are currently in preparation that would have improved impact by the inclusion of the work proposed here.

Eldridge and Barnhart will perform much of the initial data analysis, with assistance from Montana State University, and develop the occupancy models, and Sepulveda will provide expertise in data visualization and automation of the data pipeline. Barnhart and Sepulveda also will seek input from the Greater Yellowstone Coordinating Committee during development of the risk assessment map(s), and Eldridge will coordinate and oversee study design and oversee the timely completion of products, documentation, and interactions with the CDI community. Eldridge has extensive experience in project management, statistical analysis, and implementation of water-quality and toxic cyanobacteria studies. Barnhart has extensive knowledge and experience with molecular genetic tools to analyze the microbial diversity and physiology of aquatic and terrestrial systems. Sepulveda is part of the USGS Western Waters Invasive Species and Disease Research Program and has extensive experience collaborating with resource managers, including the Greater Yellowstone Network (GRYN) to synthesize and integrate vital signs data to document trends and phenological shifts within GRYN aquatic resources.

Sustainability, Outreach, and Communication

This proposed work will leverage existing resources to improve access and functionality of data by creating a novel, but proven, method for using high-throughput data streams to inform resource managers of the risk of invasive species and other hazards. These resources, along with ongoing collaborations between USGS, MBARI, NOAA, Montana State University, and Yellowstone National Park, and other stakeholder support, will ensure the sustainability and replicability of this work. Work to further develop, expand, and iteratively improve the eDNA-based risk assessment maps resulting from this study and efforts to extend the workflow to other species of management concern will continue after the CDI funding period is complete. A key outcome will be the creation of an automated pipeline for incorporating new data into existing risk assessment maps in near real time. We will achieve this by developing a flexible, user-friendly workflow using open-source software. All data sources, supporting materials, and appropriate metadata will be managed and documented in ScienceBase and in compliance with the Office of Science Quality and Integrity Instructional Memoranda on data management.

Budget Justification

The budget for this project is attached and will be used to efficiently develop the USGS products proposed here. CDI funds will primarily support USGS team members, Eldridge, Barnhart, and Sepulveda, to synthesize data, develop data visualization tools, and document the data science pipeline as a USGS Data Release. Travel expenses include funds for Elliott Barnhart to attend the 2019 CDI Workshop in Boulder, CO, and for Barnhart and Eldridge to meet with collaborators periodically at MSU (98 miles). No equipment will be purchased, but $2,000 is proposed for production of a USGS Data Release and preparation of a journal article. In addition to these costs, we propose to provide $43,800 in matching funds, which will come from various existing projects through the Water Resources and Ecosystems Mission Areas and will supplement salary cost to Barnhart, Sepulveda, and collaborators at MSU.

Timeline, FY2019

TASK	MAY-JUNE	JUNE	JULY	AUGUST	SEPTEMBER
Planning, data preparation					
Model development and refinement					
Seek input from GYCC					
Build data visualization tools					
Develop data pipeline					
Complete Data Release in ScienceBase					

References

Al-Chokhachy, R., Alder, J., Hostetler, S., Gresswell, R., and Shepard, B., 2013, Thermal controls of Yellowstone cutthroat trout and invasive fishes under climate change: Global Change Biology, v. 19, no. 10, p. 3069-3081.

Baird, D.J., and Hajibabaei, M., 2012, Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing: Molecular Ecology, v. 21, no. 8, p. 2039-2044.

Brooks, S., Gelman, A., Jones, G.L., and Meng, X.L., 2011, Handbook of Markov chain Monte Carlo: Boca Raton, FL, Chapman & Hall/CRC.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J., 2018, shiny: web application framework for R (R package version 1.2.0. ed.).

Dafforn, K.A., Johnston, E.L., Ferguson, A., Humphrey, C.L., Monk, W., Nichols, S.J., Simpson, S.L., Tulbure, M.G., and Baird, D.J., 2016, Big data opportunities and challenges for assessing multiple stressors across scales in aquatic ecosystems Marine and Freshwater Research, v. 67, no. 4, p. 393-413.

Dorazio, R., and Erickson, R.A., 2017, eDNAoccupancy: An R package for multi-scale occupancy modeling of environmental DNA data: Molecular Ecology Resources, v. 18, no. 2, p. 368-380.

Koza, J.R., Bennett, F.H., Andre, D., and Keane, M.A., 1996, Automated design of both the topology and sizing of analog electrical circuits using genetic programming, *in* Gero, J.S., and Sudweeks, F., eds., Artificial Intelligence in Design ’96: Dordrecht, Springer Netherlands, p. 151-170.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., and Hines, J.E., 2018, Occupancy applications, *in* MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., and Hines, J.E., eds., Occupancy estimation and modeling (second edition): Boston, Academic Press, p. 27-70.

National Academies of Sciences Engineering and Medicine, 2018, Future water priorities for the nation: Directions for the U.S. Geological Survey Water Mission Area: Washington, DC, The National Academies Press, 110 p.

Nichols, J.D., Bailey, L.L., O'Connell, A.F., Talancy, N.W., Grant, E.H.C., Gilbert, A.T., Annand, E.M., Husband, T.P., and Hines, J.E., 2008, Multi-scale occupancy estimation and modelling using multiple detection methods: Journal of Applied Ecology, v. 45, no. 5, p. 1321-1329.

Olson, S.A., and Norris, J.M., 2007, U.S. Geological Survey streamgaging...from the National Streamflow Information program 2005-3131 [Fact Sheet].

R Core Team, 2017, R: A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing.

Ries, K.G., Guthrie, J.G., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, StreamStats: A water resources web application.

Scholin, C.A., Birch, J., Jensen, S., Marin III, R., Massion, E., Pargett, D., Preston, C., Roman, B., and Ussler III, W., 2017, The quest to develop ecogenomic sensors: A 25-Year history of the Environmental Sample Processor (ESP) as a case study: Oceanography, v. 30, no. 4, p. 100-113.

Van den Brink, P.J., Choung, C.B., Landis, W., Mayer-Pinto, M., Pettigrove, V., Scanes, P., Smith, R., and Stauber, J., 2016, New approaches to the ecological risk assessment of multiple stressors Marine and Freshwater Research, v. 67, no. 4, p. 429-439.

SARA L. (CALDWELL) ELDRIDGE

Microbiologist, U.S. Geological Survey, WY-MT Water Science Center, Helena, MT
Phone: 406-457-5937
Email: seldridge@usgs.gov

Professional Preparation

2005-2009	M.S., Biology, Microbiology emphasis Portland State University, Portland, OR
2001-2005	B.S., Earth and Planetary Sciences (distributed minor) University of New Mexico, Albuquerque, NM
1994-1997	B.S., Zoology (major) and Biochemistry (minor) Texas State University, San Marcos, TX

Appointments and Experience

Microbiologist, Water Mission Area, U.S. Geological Survey, WY-MT Water Science Center, Helena, MT. 2018-present.

Microbiologist, Water Mission Area, U.S. Geological Survey, OR Water Science Center, Klamath Falls Field Station, Klamath Falls, OR. 2013-2018.

Hydrologist, Water Mission Area, U.S. Geological Survey, OR Water Science Center, Klamath Falls Field Station, Klamath Falls, OR. 2010-2013.

Hydrologic Technician, Water Mission Area, U.S. Geological Survey, OR Water Science Center, Klamath Falls Field Station, Klamath Falls, OR. 2009-2010.

NSF IGERT Fellow, Subsurface Biosphere, Portland State University, Portland, OR. 2005-2008

Selected Publications

Caldwell Eldridge, S.L. and Wood, T.M., Inter- and intra-annual variations of bloom-forming cyanobacteria and microcystin occurrence in Upper Klamath Lake, Oregon, based on high-throughput DNA sequencing and qPCR, *Lake and Reservoir Management, in review.*

Caldwell Eldridge, S.L., Driscoll, C., and Dreher, T.W, 2017, Using high-throughput DNA sequencing, genetic fingerprinting, and quantitative PCR as tools for monitoring bloom-forming and toxigenic cyanobacteria in Upper Klamath Lake, Oregon, 2013 and 2014: U.S. Geological Survey Scientific Investigations Report 2017-5026, 50 p., *https://pubs.er.usgs.gov/publication/sir20175026*

Caldwell Eldridge, S.L., Wherry, S.A., and Wood, T.M., 2014, Statistical analysis of the water-quality monitoring program, Upper Klamath Lake, Oregon, and optimization of the program for 2013 and beyond: U.S. Geological Survey Open-File Report 2014-1009, 82 p., *http://dx.doi.org/10.3133/ofr20141009*

Caldwell Eldridge, S.L, Wood, T.M., Echols, K.R., and Topping, B.R., 2013, Microcystins, nutrient

dynamics, and other environmental factors during blooms of non-microcystin-producing *Aphanizomenon flos-aquae* in Upper Klamath Lake, Oregon, 2009, *Lake and Reservoir Management*, 29: 68-81

Caldwell Eldridge, S.L., Wood, T.M., and Echols, K.R., 2012, Spatial and temporal dynamics of cyanotoxins and their relation to other water quality variables in Upper Klamath Lake, Oregon, 2007–09: U.S. Geological Survey Scientific Investigations Report 2012–5069, 34 p., *https://pubs.er.usgs.gov/publication/sir20125069*

Caldwell, S., Liu, Y. Ferrera, I., Beveridge, T. and Reysenbach, A.-L., 2010, *Thermocrinis minervae* sp. nov., a hydrogen- and sulfur-oxidizing, thermophilic member of the *Aquificales* from a Costa Rican terrestrial hot spring, *International Journal of Systematic and Evolutionary Microbiology*, 60: 338-343

Caldwell, S.L., Laidler, J.R., Brewer, E.A., Eberly, J.O., Sandborgh, S.C., and Colwell F.S., 2008, Anaerobic Oxidation of Methane: Mechanisms, Bioenergetics, and the Ecology of Associated Microorganisms, *Environmental Science and Technology*, 42: 18, 6791-6799

Recent Competitive Grants

Synoptic survey of cyanobacterial bloom distribution, toxicity, and associated water-quality conditions in Bighorn Canyon National Recreation Area, Montana and Wyoming. USGS / National Park Service Water Quality Partnership. 2019-2020. $75,000. Lead PI.

Evaluation of phosphorus and sediment retention through diffuse source treatment wetlands using surrogate regression modeling to improve water quality for endangered suckers in Upper Klamath Lake, Oregon. USGS / U.S. Fish and Wildlife Service Partnership. 2018-2020. $80,000. Lead PI.

ELLIOTT BARNHART

Research Hydrologist (Microbiologist), U.S. Geological Survey, Helena, MT
Adjunct Professor, Microbiology and Immunology, Montana State University
Phone: 406-853-2643
Email: epbarnhart@usgs.gov

Professional Preparation

Montana State University	Environmental Microbiology	B.S. 2009
Montana State University	Microbiology	M.Sc. 2011
Montana State University	Microbiology	Ph.D. 2014

Appointments and Experience

Adjunct Professor, Microbiology & Immunology, Montana State University,	2014-
Assistant Research Professor, Center for Biofilm Engineering	2014-
Project/Task Chief: USGS Microbial Methanogenesis of Coal Project	2014-
USGS Student Career Experience Program Research Fellow	2011-2014
Fellow, NSF Developing Global Scientists and Engineers Program, Bergen, Norway	2009
Research Assistant, Montana Bureau of Mines and Geology, Montana Tech	2008

Selected Publications

Barnhart, E.P., Davis, K., Varonka, M.S., Orem, W.H., Cunningham, A.B., Ramsay, B.D., and Fields, M.W., 2017, Enhanced coal-dependent methanogenesis coupled with algal biofuels: Potential water recycle and carbon capture: International Journal of Coal Geology, v. 171, p. 69–75.

Hodgskiss, L.H., Nagy, J**., Barnhart, E.P**., Cunningham, A.B., Fields, M.W., 2016, Cultivation of a native alga for biomass and biofuel accumulation in coal bed methane production water, Algal Research V.16, p. 63-68. doi:10.1016/j.algal.2016. 07.014

Barnhart, E.P., McClure, M.A., Johnson, K., Cleveland, S., Hunt, K.A., Fields, M.W., 2015, Potential Role of Acetyl-CoA Synthetase (acs) and Malate Dehydrogenase (mae) in the Evolution of the Acetate Switch in Bacteria and Archaea, Sci Rep., 5:12498. doi: 10.1038/srep12498.

Ritter, D., Vinson, D., **Barnhart, E.P**., Akob, D. M., Fields, M. W., Cunningham, A. B., Orem, W., McIntosh, J. C., Enhanced, 2015, Microbial Coalbed Methane Generation: A Review of Research, Commercial Activity, and Remaining Challenges, Int. J. Coal Geol. V.46, p 28-41, doi: 10.1016/j.coal.2015.04.013

Sando, S.K., Clark, M.L., Cleasby, T.E., **Barnhart, E.P**., 2015, Water-Quality Trends for Selected Sites in the Boulder River and Tenmile Creek Watersheds, Montana, Based on Data Collected During Water Years 1997-2013: U.S. Geological Survey Scientific Investigations Report 2015–5008, 46 p.

Sando, S.K., Veccia, A.V., Lorenz, D.L., **Barnhart, E.P**., 2014, Water-Quality Trends for Selected Sampling Sites in the Upper Clark Fork Basin, Montana, Water Years 1996-2010: U.S. Geological Survey Scientific Investigations Report 2013-5217

Sando, S.K., Veccia, A.V., **Barnhart, E.P**., Sando, T.R., Clark, M.L., Lorenz, D.L., 2014, Trends in major-ion constituents and properties for selected sampling sites in the Tongue and Powder River watersheds, Montana and Wyoming, based on data collected during water years 1980-2010: U.S. Geological Survey Scientific Investigations Report 2013-5179, p. 123

Barnhart, E.P., Bowen De León, K., Ramsay, B.D., Cunningham, A.B., Fields, M.W., 2013, Investigation of coal-associated bacterial and archaeal population from a diffusive microbial sampler (DMS): Int. J. Coal Geol., vol. 115 p. 64-70

Jones, E.J.P., Harris, S.H., **Barnhart, E.P**., Orem, W.H., Clark, A.C., Corum, M.D., Kirshtein, J.D., Varonka, M.S., Voytek, M.A., 2013, The effect of coal bed dewatering and partial oxidation on biogenic methane potential: Int. J. Coal Geol., vol. 115 p. 54-63

Bouskill, N.J., **Barnhart, E.P**., Galloway, T.S., Handy, R.D., Ford, T.E., 2007, Quantification of changing *Pseudomonas aeruginosa sodA*, *htpX* and *mt* gene abundance in response to trace metal toxicity: a potential in situ biomarker of environmental health: FEMS Microbiology Ecology, vol. 60, p. 276-286

ADAM SEPULVEDA

Research Zoologist
US Geological Survey, Northern Rocky Mountain Science Center
2327 University Way, Suite 2, Bozeman, MT
Phone: (406) 994-7975
email: asepulveda@usgs.gov

Professional Preparation

Postdoctoral Associate, University of Montana, April 2010 – November 2010.
Ph.D., University of Montana, Organismal Biology and Ecology, 2010.
M.S. certificate, Teton Science School, Science education, 2004.
B.A., Dartmouth College, Environmental biology and English, cum laude, 2002.

Selected publications

Sepulveda, A.J., J.J. Amberg, and E. Hanson (*in press*). Using environmental DNA to extend the window of early detection for dreissenid mussels. *Management of Biological Invasions.*

Sepulveda, A.J., P.R. Hutchins, R.L. Massengill and K.J. Dunker. 2018. Tradeoffs of a portable, field-based environmental DNA platform for detecting invasive northern pike (*Esox lucius*) in Alaska. *Management of Biological Invasions*, *9*(3), pp.253-258.

Hutchins, P.R., **A.J. Sepulveda**, R.M. Martin, and L.R. Hopper. (2018) Improved conventional PCR assay for detecting *Tetracapsuloides bryosalmonae* DNA in fish tissues. Journal of Aquatic Animal Health.

Sepulveda, A.J. 2018. Novel application of explicit dynamics occupancy models to ongoing aquatic invasions. Journal of Applied Ecology 55(2): 917-925.

Hutchins, P.R., **A.J. Sepulveda**, R.M. Martin, and L.R. Hopper. (2017) A probe-based quantitative PCR assay for detecting *Tetracapsuloides bryosalmonae* in fish tissue and environmental DNA water samples. Conservation Genetics Resources**:** 1-3

Dunker, K.J., **A.J. Sepulveda**, R.L. Massengill, J.B. Olsen, O.L. Russ, J.K. Wenburg, and A. Antonovich. 2016. Potential of environmental DNA to evaluate Northern pike (*Esox lucius*) eradication efforts: an experimental test and case study. PLoS One 11(9): e0162277. doi:10.1371/journal.pone.0162277

Sepulveda, A.J., M.T. Tercek, R. Al-Chokhachy, A.M. Ray, D.P. Thoma, B.R. Hossack, G.T. Pederson, A.W. Rodman and T. Olliff. 2015. The shifting climate portfolio of the Greater Yellowstone Area. PLoS ONE 10(12). doi:10.1371/journal.pone.0145060

Goldberg, C.S., **A.J. Sepulveda**, A. Ray, J. Baumgardt, and L. Waits. 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). Freshwater Science 32: 792-800. doi: 10.1899/13-046.1

Sepulveda, A., A. Ray, R. Al-Chokhachy, C. Muhlfeld, J. Gross, B. Gresswell and J. Kershner. 2012. An alternative approach to aquatic invasive species: lesson from cancer research and treatment. American Scientist 100: 234-239. doi: 10.1511/2012.96.234

Recent Competitive Grants

USGS stream-supergage surveillance networks for disease and invasive organisms. USGS Ecosystems Invasive Species Program. 2019. $68,358. Co-PI.

Developing protocol standards for environmental DNA monitoring of dreissenid mussels in the West. USGS Ecosystems Invasive Species Program. 2019. $72,342. Lead PI.

Predicting the spread of aquatic invasive species using remote sensing, genetics, and climate modeling. NASA Research Opportunities in Earth and Space Science. 2019 – 2021. $1,014,548. Co-PI.

Community for Data Integration (CDI)
Data Management Plan for Full Proposals

Instructions: Fill out all relevant fields of the following tables to help your team plan for your project's data management and product communication needs. For more guidance on data management plans, see the *USGS Data Management Website*, specifically the *Data Management checklist*. All products resulting from CDI projects must comply with the *Office of Science Quality and Integrity Instructional Memoranda* on data management.

Data Inputs	
Title	**Source/URL**
2018 eDNA, presence-absence via qPCR, Yellowstone River	Stored on servers maintained by NOROCK and the USGS WY-MT Water Science Center; presented in a USGS Data Release
USGS streamgage data	NWIS; https://waterdata.usgs.gov/nwis/sw
NOAA weather station data	https://www.ncdc.noaa.gov/cdo-web/

Data Processing	
Access and Sharing	Streamgage and weather input data are publicly accessible (see above), and eDNA data will be compiled and shared as a USGS Data Release. Maps will be available via a publicly accessible web service. Data will be shared between team members through USGS Cloud Hosting Services (https://support.chs.usgs.gov/display/ -CHSKB/Services).

Proposed Products I	
Title	Real-time risk assessment map enhancement
Product Type	Web Application
Description	Integration with existing USGS AIS explorer
Format	R code and Mapbox (https://www.mapbox.com/)
Data Volume Estimate	6-10 MB
Backup & Storage	https://github.com/usgs
Metadata	Metadata will be created according to USGS standards endorsed by the Federal Geographic Data Committee (FGDC).
Repository for Product	https://gagecarto.github.io/aisExplorer/
Communication Plan	Maps will be shared via the AIS Explorer website and through ongoing communications with stakeholders, in person, over phone, and email. The application will also be described in project documents, refereed publications, and scientific meetings.

Proposed Products II	
Title	Data science pipeline
Product Type	Source Code

Description	Data acquisition, data preparation, data exploration, statistical analyses, and visualization
Format	.csv and R code
Data Volume Estimate	30-50 GB
Backup & Storage	Data will consist of contributions by project participants and will be backed up and archived at contributing USGS Science Centers
Metadata	Metadata will be created according to USGS standards endorsed by the Federal Geographic Data Committee (FGDC).
Repository for Product	https://github.com/usgs
Communication Plan	Product will be shared via ongoing communications with stakeholders, in person, over phone, and email. The application will also be described in project documents, refereed publications, and scientific meetings.

Proposed Products III	
Product Type	Data Release
Description	Data compilation (non-NWIS) and pipeline documentation
Format	.pdf files
Data Volume Estimate	30-50 GB
Backup & Storage	Data will consist of contributions by project participants and will be backed up and archived at contributing USGS Science Centers
Metadata	Metadata will be created according to USGS standards endorsed by the Federal Geographic Data Committee (FGDC).
Repository for Product	USGS ScienceBase, https://www.sciencebase.gov/catalog/
Communication Plan	Product will be shared via ongoing communications with stakeholders, in person, over phone, and email. The application will also be described in project documents, refereed publications, and scientific meetings.

CDI Product Type Vocabulary

Data Release: A formal USGS data release that will go through FSP review and approval
Mobile Application: Interactive application built specifically for a mobile device
Presentation: Slides, video, or other presentation media
Publication: Peer-reviewed publication (USGS or external journal publication)
Software: Executable or compiled code that can be downloaded
Source Code: A code repository for the project's source code
Web Application: Interactive application that runs on a web browser
Web Link: Project webpage, wiki page, white paper, or online resources that do not fit other categories
Web Service: A service endpoint URL where your service can be accessed by a client application

Community For Data Integration (CDI) RFP BUDGET FORM

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
GRAND TOTAL:	$49,979	$43,800
Do not edit the rows above this line.		
1. PERSONNEL (SALARIES including benefits):		
Personnel:		
Sara Eldridge, 240hrs at $47/hr	$12,900	
Elliott Barnhart, 40hrs at $54/hr	$1,300	$1,300
Adam Sepulveda, 200hrs at $65/hr	$7,500	$7,500
Contract Personnel :		
University Collaborators, 750hrs at $60/hr	$10,000	$35,000
Total Salaries:	$31,700	$43,800
2. TRAVEL EXPENSES:		
Trip 1 (CDI conference, 7 days, 1 travelers)		
Per Diem:	$1,529	
Transportation (Airfare + Mileage/Shuttle):	$600	
Other expenses (e.g. registration fees):		
Trip 2 (Travel to MSU, 6 days, 1 travelers)		
Per Diem:	$300	
Transportation (Airfare + Mileage/Shuttle):		
Other expenses (e.g. registration fees):		
Total Travel Expenses:	$2,429	$0
3. OTHER DIRECT COSTS: (itemize)		
Equipment (inc. software, hardware, purchases/rentals):		
Publication Costs:	$2,000	
Office supplies:		
Training:		
Other expenses (specify):		
Total Other Direct Costs:	$2,000	$0
Total Direct Costs:	$36,129	$43,800
Indirect Cost:	$13,850	